

(19)



Europäisches Patentamt

European Patent Office

Office européen des brevets



(11)

EP 0 732 401 A2

(12)

EUROPEAN PATENT APPLICATION

(43) Date of publication:
18.09.1996 Bulletin 1996/38

(51) Int. Cl.⁶: **C12N 15/27**, C07K 14/53,
C12N 5/10, A61K 38/19

(21) Application number: **95119320.0**

(22) Date of filing: **07.08.1990**

(84) Designated Contracting States:
AT BE CH DE DK ES FR GB IT LI LU NL SE

(30) Priority: **08.08.1989 US 390901**
28.12.1989 US 457196
29.06.1990 US 546114

(62) Application number of the earlier application in
accordance with Art. 76 EPC: **90913091.6**

(71) Applicant: **GENETICS INSTITUTE, INC.**
Cambridge, Massachusetts 02140 (US)

(72) Inventors:
• **Turner, Katherine**
Melrose, Massachusetts 02176 (US)

• **Clark, Steven C.**
Winchester, Massachusetts 01890 (US)
• **Gesner, Thomas G.**
Kensington, CA 94707 (US)
• **Hewick, Rodney M.**
Boston, MA 02116 (US)

(74) Representative: **VOSSIUS & PARTNER**
Postfach 86 07 67
81634 München (DE)

Remarks:

This application was filed on 07 - 12 - 1995 as a
divisional application to the application mentioned
under INID code 62.

(54) **Megakaryocyte stimulating factor (meg-CSF)**

(57) A novel human megakaryocytopoietic factor
capable of stimulating the growth and development of
colonies of megakaryocytes is provided, including pro-
cedures for its purification and use as a pharmaceutical
agent.

EP 0 732 401 A2

Description

The present invention relates generally to a novel protein factor which is important in regulating the human hematopoietic system. More specifically the invention discloses a novel protein factor that stimulates megakaryocytic colony formation and the differentiation or maturation of megakaryocyte progenitors. Also provided are processes for obtaining the factor in homogeneous form and producing it by recombinant genetic engineering techniques.

Megakaryocytes are the hematopoietic cells, largely found in the bone marrow, but also in peripheral blood and perhaps other tissues as well, which produce platelets (also known as thrombocytes) and subsequently release them into circulation. Megakaryocytes, like all of the hematopoietic cells of the human hematopoietic system, ultimately derive from a primitive stem cell after passing through a complex pathway comprising many cellular divisions and considerable differentiation and maturation.

The platelets derived from these megakaryocytic cells are critical for initiating blood clot formation at the site of injury. Platelets also release growth factors at the site of clot formation that speed the process of wound healing and may serve other functions. However, in patients suffering from depressed levels of platelets (thrombocytopenia) the inability to form clots is the most immediate and serious consequence, a potentially fatal complication of many therapies for cancer. Such cancer patients are generally treated for this problem with platelet transfusions. Other patients frequently requiring platelet transfusions are those undergoing bone marrow transplantation or patients with aplastic anemia.

Platelets for such procedures are obtained by plateletpheresis from normal donors. Like most human blood products, platelets for transfusion have a relatively short shelf-life and also expose the patients to considerable risk of exposure to dangerous viruses, such as the human immunodeficiency virus (HIV).

Clearly the ability to stimulate endogenous platelet formation in thrombocytopenic patients with a concomitant reduction in their dependence on platelet transfusion would be of great benefit. In addition the ability to correct or prevent thrombocytopenia in patients undergoing radiation therapy or chemotherapy for cancer would make such treatments safer and possibly permit increases in the intensity of the therapy thereby yielding greater anti-cancer effects.

For these reasons considerable research has been devoted to the identification and purification of factors involved in the regulation of megakaryocyte and platelet production. Although there is considerable controversy, the factors regulating the growth and differentiation of hematopoietic cells into mature megakaryocytes and the subsequent production of platelets by these cells are believed to fall into two classes: (1) megakaryocyte colony-stimulating factors (meg-CSFs) which support the proliferation and differentiation of megakaryocytic progenitors in culture, and (2) thrombopoietic (TPO) factors which support the differentiation and maturation of megakaryocytes resulting in the production and release of platelets. [See, e.g., E. Mazur, Exp. Hematol., 15:340-350 (1987).]

Either class of factors is defined by bioassay. Factors with meg-CSF activity support megakaryocyte colony formation, while factors with TPO activity elicit an elevation in the numbers of circulating platelets when administered to animals. It is not clear how many species of factors exist that have either or both of these activities. For example, human IL-3 supports human megakaryocyte colony formation and, at least in monkeys, also frequently elicits an elevation in platelet count. However, IL-3 influences hematopoietic cell development in all of the hematopoietic lineages and can be distinguished from specific regulators of megakaryocytopoiesis and platelet formation which interact selectively with cells of the megakaryocytic lineage.

From the studies reported to date, it is not clear whether activities identified as meg-CSF also have TPO activity or vice versa. Many different reports in the literature describe factors which interact with cells of the megakaryocytic lineage. Several putative meg-CSF compositions have been derived from serum [See, e.g., R. Hoffman et al, J. Clin. Invest., 75:1174-1182 (1985); J. E. Straneva et al, Exp. Hematol., 15:657-663 (1987); E. Mazur et al, Exp. Hematol., 13:1164-1172 (1985)]. A larger number of reports of a TPO factor are in the art. [See, e.g., T. P. McDonald, Exp. Hematol., 16:201-205 (1988); T. P. McDonald et al, Biochem. Med. Metab. Biol., 37:335-343 (1987); T. Tayrien et al, J. Biol. Chem., 262: 3262-3268 (1987) and others].

Although there have been numerous additional reports tentatively identifying such regulatory factors, the biochemical and biological identification and characterization of these factors has been hampered by the small quantities of the naturally occurring factors available from natural sources, e.g., blood and urine. At present there is no identification of a single purified factor useful as a meg-CSF or TPO for pharmaceutical use in replacing serum-derived products or platelets.

There remains a need in the art for additional proteins purified from their natural sources or otherwise produced in homogeneous form, which are capable of stimulating or enhancing the production of platelets *in vivo*, to replace presently employed platelet transfusions.

In one aspect the present invention provides a novel human megakaryocytopoietic factor (meg-CSF) which is substantially free from other human proteins. This protein may be produced by recombinant genetic engineering techniques. It may also be purified from cell sources producing the factor naturally or upon induction with other factors. meg-CSF may also be synthesized by chemical techniques, or a combination of the above-listed techniques.

The meg-CSF of the present invention has been found to stimulate the growth and development of colonies consisting of intermediate and large size megakaryocytes in an assay using murine bone marrow target cells. meg-CSF displays biological activity in this assay of greater than 5×10^7 dilution units per milligram of protein. meg-CSF has also displayed activity in an assay using human cells, as described in Example 8 below.

Active meg-CSF has an apparent molecular weight of approximately 28-38 kd as determined by sodium dodecyl sulfate polyacrylamide gel electrophoresis under non-reducing conditions. meg-CSF has an apparent molecular weight of approximately 20-27 kd as determined by sodium dodecyl sulfate polyacrylamide gel electrophoresis under reducing conditions.

The active approximately 28-38 kd meg-CSF is further characterized by comprising all or a portion of the sequence of Tables I through IV. Meg-CSF is also characterized by comprising at least one of the same or substantially the same four amino acid sequences or fragments thereof, recited below as sequences (a) through (d).

Another aspect of the present invention is a DNA sequence that encodes the expression of a human meg-CSF protein. This DNA sequence may include an isolated DNA sequence that encodes the expression of a human meg-CSF protein as described above. The DNA sequence coding for all or a portion of the meg-CSF protein is characterized as comprising the same or substantially the same nucleotide sequence in Table I, II, or III or fragments thereof. This DNA sequence may include additional coding sequence. The DNA sequence may also include 5' and 3' human non-coding sequences flanking the meg-CSF coding sequence. The DNA sequence may also encode an amino terminal signal peptide. Table I illustrates a putative partial genomic sequence beginning at the extreme 5' border of the genomic clone and ending at the Bgl II site in the intron occurring between Exons II and III. Table I contains sequence encoding the open reading frames of Exon I and Exon II. It may also contain the N-terminal Exon and any other additional Exons which are contained in the genomic sequence 5' of Exon I which are present in the full length Meg cDNA. Table II illustrates the other known Exon III forming a partial genomic coding sequence of human meg-CSF isolated from human urine and expressed in COS-1 cells. Table III illustrates a putative partial CDNA coding sequence containing Exons I through III derived from the genomic sequence.

It is understood that the DNA sequence of this invention may encode a biologically active human meg-CSF protein and may also comprise DNA sequences capable of hybridizing under appropriate conditions, or which would be capable of hybridizing under said conditions, but for the degeneracy of the genetic code, to an isolated DNA sequence of Table I, II, or III. Thus, the DNA sequence of this invention may include or contain modifications in the non-coding sequences, signal sequences or coding sequences based on allelic variation, species variation or deliberate modification.

Still a further aspect of the present invention is a process for isolating and purifying the meg-CSF composition of the present invention or a fragment thereof from human urine. This purification process provided by the present invention involves the steps of concentrating the urine; subjecting it to anion exchange column chromatography; followed by cation exchange column chromatography; subjecting the resulting materials to lectin affinity chromatography followed by cation exchange fine performance liquid chromatography (FPLC) and three elutions through reverse phase high pressure liquid chromatography (HPLC) using different solvent solutions for each HPLC run.

A further aspect of the present invention is homogeneous meg-CSF, purified from urine or produced via recombinant or synthetic techniques, which is characterized by a specific activity in the murine fibrin clot assay of greater than 5×10^7 dilution units/mg.

Also provided by the present invention is a recombinant DNA molecule comprising vector DNA and an DNA sequence encoding human meg-CSF. The DNA molecule provides the meg-CSF DNA in operative association with a regulatory sequence capable of directing the replication and expression of meg-CSF in a selected host cell. Host cells transformed with such DNA molecules for use in expressing recombinant meg-CSF protein are also provided by the present invention.

The DNA molecules and transformed cells of the invention are employed in another aspect, a novel process for producing recombinant human meg-CSF protein, or peptide fragments thereof. In this process a cell line transformed with a DNA sequence encoding expression of meg-CSF protein or a fragment thereof (or a recombinant DNA molecule as described above) in operative association with a suitable regulatory or expression control sequence capable of controlling expression of the protein is cultured under appropriate conditions permitting expression of the recombinant DNA. The expressed meg-CSF protein is then harvested from the host cell, cell lysate or culture medium by suitable conventional means. The conditioned medium may be processed through the same purification steps or modifications thereof as used to isolate the meg-CSF from urine. This claimed process may employ a number of known cells as host cells for expression of the protein. Presently preferred cell lines for producing meg-CSF are mammalian cell lines and bacterial cells.

As still a further aspect of the present invention, there is provided recombinant meg-CSF protein. This protein is substantially free from other human proteinaceous materials and comprising a DNA sequence encoding one or more of the peptide fragments or sequences described herein. The meg-CSF protein of this invention is also characterized by containing one or more of the physical, biochemical, pharmacological or biological activities described herein.

Another aspect of this invention provides pharmaceutical compositions containing a therapeutically effective amount of homogeneous or recombinant meg-CSF or an effective amount of one or more active peptide fragments thereof. These pharmaceutical compositions may be employed in methods for treating disease states or disorders characterized by a deficiency of platelets. Thus the meg-CSF composition of the present invention or pharmaceutically effective fragments thereof may be employed in the treatment of a plastic anemias resulting from chemotherapy or thrombocytopenia. meg-CSF may be used as an adjunctive therapy for bone marrow transplant patients.

A further aspect of the invention, therefore, is a method for treating these and other pathological states resulting from a deficiency of platelets by administering to a patient a therapeutically effective amount of meg-CSF or one or more peptide fragments thereof in a suitable pharmaceutical carrier. These therapeutic methods may include administering simultaneously or sequentially with meg-CSF or one or more peptide fragments thereof an effective amount of at least one other TPO-like factor, a cytokine, hematopoietin, interleukin, growth factor, or antibody.

Still another aspect of the present invention are antibodies directed against human meg-CSF or a fragment thereof. As part of this aspect, therefore, the invention claims cell lines capable of secreting such antibodies and methods for their production and use in diagnostic or therapeutic procedures.

Other aspects and advantages of the present invention will be apparent upon consideration of the following detailed description of preferred embodiments thereof.

Fig. 1 illustrates a restriction map illustrating the position of restriction endonuclease enzymes of the 18.2 kb genomic clone containing meg-CSF. The lower portion of the Figure illustrates the genomic regions in which Exons I, II and III are located, and shows their relative positions in the restriction map.

The novel human megakaryocyte colony stimulating factor, meg-CSF, provided by the present invention is a homogeneous protein or proteinaceous composition substantially free of association with other human proteinaceous materials. This protein can be produced via recombinant techniques to enable large quantity production of pure, active meg-CSF useful for therapeutic applications. Alternatively this protein may be obtained as a homogeneous protein purified from human urine or a mammalian cell line secreting or expressing it. Further meg-CSF or active fragments thereof may be chemically synthesized.

meg-CSF of the present invention is characterized by one or more of the following biochemical and biological properties:

(1) The composition of the present invention has an apparent molecular weight of approximately 28-38 kd as determined by 12% sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) under non-reducing conditions and by murine fibrin clot megakaryocyte colony formation bioassay;

(2) The composition of the present invention has an apparent molecular weight of approximately 20-27 kd as determined by 12% SDS-PAGE under reducing conditions using a variety of reducing agents, e.g., beta-mercaptoethanol or dithiothreitol;

(3) The composition of the present invention has a specific activity in the murine fibrin clot megakaryocyte colony formation assay of greater than approximately 5×10^7 dilution units/mg protein.

(4) The meg-CSF composition of the present invention contains one or more of the same or substantially the same amino acid sequences or fragments thereof:

(a) Ser Arg Cys Phe Glu Ser Phe Glu Arg

(b) Arg Val Cys Thr Ala Glu Leu Ser Cys Lys Gly (Arg)

(c) Lys Ala Pro Pro Pro (X) Gly Ala Ser Gln Thr Ile Lys

(d) Lys Tyr Asp Lys Cys Cys Pro Asp Tyr Glu Ser Phe Cys Ala Glu Val His Asn Pro

(e) an amino acid sequence contained in Table I, IA, IB, IC, II, or III below.

These sequences or fragments of these sequences may also have biological or physiological activity similar to that of the complete meg-CSF protein. In the sequences, (X) indicates that the residue is not yet absolutely identified, but may be Ser or Thr; and () indicates tentative identification of a residue. The sequences identified above as (a) through (d) were originally determined from purified material from step 8, the third HPLC purification step of the purification procedure, omitting step 7. The same sequences also have been obtained from the material, when purified through all eight steps. The DNA and amino acid sequences of Tables I through III are discussed in detail below.

(5) The meg-CSF composition of the present invention is capable of binding SP-Zeta Prep under acidic conditions of pH 4.5.

(6) The meg-CSF composition of the present invention is capable of binding to Wheat Germ-Sepharose and Concanavalin-A Sepharose.

(7) The meg-CSF composition of the present invention elutes between 23-33% acetonitrile on a reverse-phase C4 HPLC column using a solvent system of trifluoroacetic acid (TFA) and acetonitrile.

(8) The meg-CSF composition of the present invention elutes between 6-15% n-propanol on a reverse-phase C18 HPLC column using a solvent system of pyridine, acetic acid and n-propanol.

(9) The meg-CSF composition of the present invention elutes between 27-37% acetonitrile on a reverse-phase C4 HPLC column using a solvent system of heptafluorobutyric acid (HFBA) and acetonitrile.

The biological activity of the meg-CSF composition of the present invention is demonstrated by its ability to stimulate the growth and development of colonies consisting of intermediate and large size megakaryocytes in culture. In the murine fibrin clot megakaryocyte colony formation assay, the meg-CSF composition of the present invention stimulates colonies of an average of 3-6 megakaryocytes. In the murine agar meg-CSF assay, the meg-CSF composition of the present invention stimulates colonies of megakaryocytes. The meg-CSF composition of the present invention has inconsistently shown activity in the human plasma clot megakaryocyte colony formation assay.

meg-CSF was originally detected in the urine of human patients with bone marrow transplants. These patients demonstrate an enhanced level of meg-CSF activity. Human meg-CSF was initially purified from this human urine by a sequence of purification steps and techniques specifically described in Example 1 below. However, this factor may also be purified from other sources, e.g., human cell lines, or produced via recombinant means from those cell lines.

The purification techniques employed in obtaining meg-CSF from the human urine comprises the following steps. The purification steps include concentrating pooled bone marrow transplant patient urine through an Amicon YM-10 filter. The concentrated urine is passed through an anion exchange chromatographic column and the flow-through is bound onto a cation exchange chromatographic column. The urinary protein eluate was then subjected to pooling, dialyzing and heating and applying it to a lectin affinity chromatographic column. This eluate is then dialyzed and applied to a cation exchange FPLC column. Finally this eluate is applied through three cycles of reverse phase HPLC using different solvent systems.

Batches with the highest levels of meg-CSF in the murine fibrin clot assay, described below, were selected for further purification at the semi-preparative scale (between 30 and 100 liters urine equivalent) to maximize recovery and yield.

Thus the homogeneous meg-CSF may be obtained by applying the above purification procedures, which are described in detail in Example 1, to human urine or other sources of human meg-CSF, e.g., activated peripheral blood leukocytes and human placenta. Other tissue sources and cell lines such as C10-MJ2 (an HTLV1-transformed T cell line) and HEK (primary human embryonic kidney cells) may also be sources of this protein. Procedures for culturing a cell source which may be found to produce meg-CSF are known to those of skill in the art.

meg-CSF or one or more peptide fragments thereof may also be produced via recombinant techniques. To obtain the genomic DNA and cDNA sequences for meg-CSF or one or more fragments thereof, tryptic digests of the purified, sequenced polypeptide were prepared, i.e. the tryptics identified as (a) through (d) above, by conventional techniques.

As described in detail in Example 5, below oligonucleotide probes were synthesized using the genetic code to predict all possible sequences that encode the amino acid sequences of the tryptic fragments or the above-identified amino terminal sequence of meg-CSF. The probes were employed to screen a human placenta lambda phage DNA library (a human genomic library). One of the probes hybridized to an 18.2 kb genomic DNA insert. A restriction map of this insert is illustrated in the top portion of Fig. 1. The lower portion of the Figure indicates the 5 kb region of the restriction map containing genomic sequences encoding three putative Exons I, II and III. The genomic sequence was obtained in three sections; NotI to EcoRI, EcoRI to BglII and BglII to BglII. The sequences were connected together at these sites to generate one long continuous stretch of genomic sequence reported in Table I. However it is possible that at these junctions there might be a small deletion which is not present in Table I.

The sequence of Table I is a partial sequence of the genomic meg-CSF clone beginning at the extreme 5' border of the clone (adjacent to the 5' Not I site of Fig. 1). Nucleotides indicated by -1 to -64 are polylinker nucleotides. The first nucleotide of the genomic sequence is indicated as n1 in the sequence. The Table I sequence ends at the second BglII site in the intron sequence between Exon II and Exon III. This sequence contains approximately 7505 nucleotides. Four enzymes, KpnI, PvuII, EcoRI, and SnaBI have the ability to cut this sequence and NotI present in the polylinker of the cloning site, adjacent to the 5' border of the genomic sequence can also cut this sequence. Their expected cutting sites are indicated in Table I at approximately the sites marked by asterisks. It is postulated that this sequence contains at least two Exons but the location of the 5' border of Exon I and the presence of any additional Exons has not been positively determined. The amino acid sequence of Exon I is not positively determined. Further the proper reading frame of the amino acid sequence of any additional Exons contained in this sequence is not positively determined. In the absence of positive identification of the proper reading frame, Tables IA, IB, and IC contain putative amino acid sequences for the reading frames beginning with the first genomic nucleotide of the genomic clone Table I (reading frame A), with the second nucleotide of Table I (reading frame B) and the third nucleotide of Table I (reading frame C), respectively. All three reading frames contain a potential N-terminal Met-containing open reading frame sequence identified by computer as a likely Exon candidate containing a secretory leader sequence. One such sequence spans nucleotide #4897 through 5073 (underlined) in Table IA. The amino acid sequence from nucleotide #5742 to #5961 of Table I (in Reading Frame A) has been determined to correspond to the coding region of Exon I. Therefore, Table IA includes

Exon I (underlined), and may be the correct reading frame sequence, assuming that no nucleotides were deleted or repeated incorrectly in the sequence. The amino acids in Tables IB and IC corresponding to nucleotides #5742 to #5961 have been deleted.

Exon II spans nucleotides #7340 through #7459 in Reading Frame B. Other potential open reading frame sequences include in Reading Frame B, a sequence spanning nucleotide #4202 through #4465 and a sequence spanning nucleotide #4466 through #4915. In Reading Frame C, one such potential sequence spans nucleotide #2559 to #2732.

The sequences of Table I through III were obtained from partial genomic clones of human meg-CSF and contain Exons I, II and III obtained as described below. The region which hybridized to the probe was subcloned, sequenced and is illustrated in Table I, with the open reading frame labeled Exon II which is found on the BglII fragment appearing directly above the bar graph labelled II in Fig. 1.

The 18.2 kb genomic DNA fragment also hybridized to two additional probes. The hybridizing regions were individually subcloned, sequenced and shown to contain the sequences illustrated in Table I and II, respectively, as Exons I and III within open reading frames. All four tryptic sequences are present in the three Exons.

Two of the tryptics overlapped intron/Exon junctions and define the borders of Exon II. The resulting putative cDNA sequence and predicted amino acid sequence (three letter code) of three Exons (I, II, III) in a single reading frame containing the meg-CSF partial cDNA sequence are reported in Table III below. The partial sequence contains 182 amino acids and 546 nucleotides, containing all four of the tryptic sequences above. The 5' and 3' borders of this cDNA are not precisely identified, indicating that an N-terminal Met-containing Exon is presently unidentified, as well as a possible extra Exon 3' to Exon III. It is presently speculated that a possible site for the 5' border of Exon I (i.e., where the N-terminal Met-containing Exon or the 5'-adjacent Exon would splice) occurs in the underlined sequences located in Tables IA, IB, or IC above (other than the underlined sequence of Exons I and II).

TABLE I

5	Partial genomic clone 5' NotI to 2nd Bgl II site, containing Exon I			
	NotI			
	*			
10	AATTCGAGCT	CCACCGCGGT	GGCGGCCGCG	AGCTCTAATA -64
	Sal I n1			
	* *			
15	CGACTCACTA	TAGGGCGTCG	ACTCGATCTT	TTACTCTGA 16
	Nco I			
	*			
20	AGGACTTTCT	CTACTCCTTT	AGACCATGGG	CAGAAATGTA 56
	SnaBI			
	*			
	CACATTATTG	GTCTACGTAG	ACAGACAAAT	TTGTAATCTC 96
25	TGAACTATAA	TTTCAAATTT	CCAGGAGAAG	AAAACATATT 136
	PvuI			
	*			
30	GGCTCGGGTT	GTTCAAGTTC	CAATTCCTAA	TCCTATCAGC 176
	TGTGGCTGAT	GTGGAAGAT	AACATATATA	ACCAGAGCTA 216
	AAAGGAAAAT	AGCCCGATAG	AAAGAGAAGT	AGTTCCCAGA 256
35	AAAGGGGGAT	TATTTAGAGA	TGAGCAGATA	GCCTCGAAAT 296
	TGTCTTTACA	TATTTAGTCC	TGTGAAATAA	GCACTGTAGA 336
	AAGAAAATAG	ATGCTTATAA	TTCTAGTCTT	AAAAAAGTCA 376
40	CAATCTCACA	AACCTATGTA	TACATGTAGA	GTAAAGGAGT 416
	AAAATTTAGT	ATAATGATTA	CAAATTAAGT	GATGGTTTTG 456
45	ATCAAAAGGG	AGCTGGATGT	AGTGGAACA	TGACGGAAC T 496
	TAAAATGCAA	AGACATGGGT	TCAAATCTTG	GTTCTGGCCT 536
50	CTAGTCTTTA	TAGATCCTTG	GATGGGTCAC	ACATTCTTCA 576
	TTTGTGGGTC	TTCATCTTCT	TATTTGTAAA	TGTGGATGAT 616
55				

EP 0 732 401 A2

	TATATTTAAC TTTAAGACAT TTTGTGAGAA TTAAATGCAG	656
5	TACCTACCAA ATGAGTAGTA GACAGCAGAA TCCAGCCTAC	696
	AACCATGTGG TTTACTGAAC ATGTTTTCTT TTTTTTCTTT	736
	TCTTTTTTTTT TTTTGTGAGAC TGAGTTTTGC TCTTGTGACC	776
10	CAAGCTGGTG TGCAGGTGGCG TGATCTCAGC TCACTGCAAC	816
	CTCTGCCTCT CAGGTTCAAG TGATTCTCCT GCCTTAGCCT	856
15	CCCGAGTAAC TGGAATTACA GGCGCCTGCC ACCACACCCG	896
	GCTAATTTTT TGTATTTTTTA GTAGAGAGGG AGTTTCACCA	936
	TGTTGGTCAG GCTGGTCTCA AACTCCTGAC CTCAGGTGAT	976
20	CTGCCCCACCT CAGCCTCCCA AAGTGCTGGG ATTGCAGGGG	1016
	TGAGCCACTG CACCCGGCGT GAGCATGTTT TAAACATATA	1056
25	ACATTGGTGA AACAACTGGG CTGTATGTTG GCAATAGACT	1096
	AACATATAGG TGTGGTTGT CCTTAGGGAA ATCATAGTCT	1136
	GTTGGGGGTG TGAGACAGAT GAAGAGATAT TTTAGTAAAA	1176
30	CGTAGTACAA TTGTTTACTA TAATTCTAGA TGTGATTGGA	1216
	ATTTGATCTG GAAAGGTTTC AGGGAGGTGC TAGGACATGA	1256
35	GCGAAGCCTT ATGTTGGTGG GTAAGATTTT ACACCGATAC	1296
	ATTCCGTATG AAAGTGGGGT AAAGCCTTGG GCTGAGCTGA	1336
	CTTAGCTCTG CAATGGTGAT GGTTACACATT TTCATTGTGG	1376
40	GAAGGAAACA AAATTAGTAA TTTAAAAGAC CTTTTCCAAA	1416
	CATGCTTATC TTAAAGGAAG GATCTAAGGA AAATATCAAT	1456
45	ACATTTTTTA AGCAGATTTT TAAAATAATA AAGGAAGCTA	1496
	GAAAATCCAG TGTCAATAAT CTACCTGTAG GTAAGTGAAG	1536
50	GCATGTTTTT TGTGCTCAGA GAAAATGAAA CTAAGTCAAGA	1576

55

EP 0 732 401 A2

	AATGTATATT TGATATAATG GTATAAGTAT ATAAAATGTA	1616
5	TGCCACAAGC TAACACTAGG TAAATTTCAA AGCTACCTTT	1656
	GCTAGATTTA ACTGGAAAGC AGAAGAAATA AATGATACCA	1696
	GTATGTGAAT ATTTTAAAGT TGCTGCTTCA ATATACTAGT	1736
10	CAAAGTCTGC CAGAGCTATT TACCATCTCA GGACAATTTG	1776
	TTCACCTCAC TTAGAAATTA GAAATGTGCT CTGGCCAATA	1816
15	CATCTAAATT TATTTTGTGG TAGTCATTTT GCTTAATGGA	1856
	AGTGTTTTCC AGATGTTGCC TAAGTCTAGT CGTCTGGGCC	1896
	CTTTTCCAAT AAATGAGGAA AGTTTGATTT CATAGGTTGT	1936
20	CACTGTTGAT TTTGTCTAAC CTTTGGACTA ATTGGTTCAT	1976
	CCCACTGTAT TTGCACTGAT ATATAAGACT CCCAGGACGT	2016
25	GGGATAAACT CATCTATCCT TTACGGGTAA GTGTCAAAGT	2056
	TAACTTGCTT CATAGAATTA AATGTGTTTT CATTAGAGGT	2096
	GTTTGAAAAA AAATGTGTAA AGATAGTAGT TGGAAATTTT	2136
30	TGAAAAGGAT TATGTTTATG CAATACACCT GTTGAAGCC	2176
	TTTGAATTTA TATTGAGAAT TAAGAAAAAG TTGGAACAGT	2216
35	AACTCCATGA TGCTTATTAA ATTACATTTT TGTGACACAG	2256
	GTTATTATTT TCCTTAAGAT AAAATTTTAA CCTTGCACTG	2296
	TTAAGTACAT GCCATACTTT GGCTAGAGTT TTAAGATAAG	2336
40	TCTATTCTAT ATTGGAGATT TCTTAAGAAT TAAAACTGTC	2376
	AATAACGTGG CTGAGGGTGA CTTGATTTGT TCTGCATAAG	2416
45	ATTAAGTCTA ATGGCCAAAT ATTTTCTATA AAATTAAAGC	2456
	TGATTTAAAG TCTTGTTTCA AGAATGGGAT GGTTTTATTA	2496
50	TTGTTACATT TTAAAGTTA CTGAAATGTG TATAATGCAA	2536

55

	GCCTAAGTTA GTGGTGAGAT GAAAGAGCTG TTTTCTGATA	2576
5	KpnI * KpnI *	
	CTTTTATTTT ATTTTCAGCA AGGGTACCTA CGGTACCTGA	2616
	AAACAACGAT GGCATGGAAA ACACTTCCCA TTTACCTGTT	2656
10	GTTGCTGCTG TCTGTTTTTCG TGATTTCAGCA AGTTTCATCT	2696
	CAAGGTAGCT TAACCATCGA ACATACTTTT ATTTAACAAC	2736
15	TATTGCTAAT CATTCAGTCT TGATTTTTTAT AACAAACGGAA	2776
	ATATATTTCT AAAAATTTAT ATTTGCTTGA GTTTAATATA	2816
	CTGTACAACC TTAAAATAAT ATATAACATG TAGCCTGTTT	2856
20	Eco RI *	
	GTAAGTGCTT TGGAATTCTC ATCTTTTTTCT CCTTTTGAGA	2896
25	GTTTTGAGCT TGTATAAAGA ATAATTTTGG TACTTGTTAT	2936
	TTCTATGGTT CTCTCGAAAG TTTAAAAAGT GCAGACAGTT	2976
	CATCATAAAA TTATGCTGGA GCTATATAGT ATCATAAAAA	3016
30	TAACTAAAAA AGAACAATAT TTAGATAACA TGTAATTTTT	3056
	GCCTATATGT TGAACAATTT GTTTTTTAAA ATCAAATAAA	3096
35	TTATTGATAA ATAACCTTATT TTGGCATCAT TTAGACATAG	3136
	ATACCATATT ATAATAAACT ATAGTGTTTC AAATGGCATT	3176
	TTATTAGTCT TGAAAATGTT CTCTTTTTTCT TACAAGAACT	3216
40	ATTTATCCAG GTATGGACAT ATACAGATGT GACTGTCATT	3256
	TTGTGTTATA ACATTAAACT TTTCACAGTT CTCCCAAAC	3296
45	AGGTCCTGGA TGAATAGTTC TTGTTACTCA TTTTLAGAGA	3336
	CTGTTAAAGT ACATTCAGTG AAATCACATT AGTAAAAGAT	3376
	GCTAGCATGC CATCTAGGGG GATTAATAGT AACAAATGCCA	3416
50	AGCTTTGGAT TTTTCTTTTC TTTTCTTTTC CTTTGTTC	3456

55

EP 0 732 401 A2

	CCTTTGATAA	GCAAAATCTG	AGAGAATAAA	AATCAAGATT	3496
5	CATGACAGTT	ATGATGAAAT	TATGTTTCTA	AAGTAAACAT	3536
	TTCCAGTAAA	ATACGAGATT	CTTATGAAAC	CTGAACATGA	3576
	GTGGTAACTG	TCTGCATAGG	CATAAGTTGC	AGAATTGTTT	3616
10	AGATAAGAAA	AGACAGGAAA	ACACATGAAA	GCAAATGTGA	3656
	ATATTCAATA	AGAATGATGA	CTACTCCAGT	ATCTCCAGAC	3696
15	CCTTCGGCTT	TCTCGTAACA	CTATGACAAG	G TTCACAACA	3736
	CTGGGGGCAC	TTTCTAAACT	GCCTTTTCCT	CTGTGATACA	3776
	ATTGGTTGTT	CACTAAAACA	GTGTTACTTT	CATTTTAATT	3816
20	GTGATTAAAT	AAATCAAATT	AAAATTAATG	GGGCTGGGTG	3856
	TGGTGGCTCA	CACCTGTAAT	CCCAGCATTC	TGGGAGGCCT	3896
25	AGGGAGACAG	ATCACCTGAG	GTCTGGAGTT	CAAGAGCAGC	3936
	CTGGGCAACA	TGGCAAAATC	TTGTCTCTAT	ACAAAATACA	3976
	AAAAAATTAG	CCAGGTGTGG	TAGTGTGTGC	CTGTAATCCC	4016
30	AGCTATTCGG	AGGCTGAGGC	AAGAGGATAT	CTTGAATCCA	4056
	GGAGGTGGAG	GTTGCATTAA	GCTGAGCTTG	CATCATGGCA	4096
35	CTCCAGCCTG	GGCAACAGAG	TGAGACTCTG	TCTCAAAAAT	4136
	AAATAAATAA	ATAAATAAAA	TTAATGGTTA	CAATTAATAG	4176
	CAATAAAAGT	AGGACAGACA	CCTAATCTAT	GAAAGTAAGC	4216
40	TTTTCTCTGTA	AGGATGATTT	CCTCTTTTTT	TTTTTTCTTT	4256
	TGAGACAGGG	TCTTGCTTTG	TCTCCCAGTC	TGGAGTGCAG	4296
45	TGGCACAATC	TTGGCTCACT	GCAACCTCCA	CCTACGGGGT	4336
	TCAAGCGATT	CTCCTGTCTC	AGCCTCCCGA	G TAGCTGGGA	4376
50	TTACAGGCCC	CTGCCACCAT	GCCCGGTAA	TTTTTGTATT	4416

55

	TTTAGTAAAG ACAGGTTTTA ACCATGTTGG CCAGGCTGGT	4456
5	CTCACCTGAC CTCAGGTGAT CCGCCCGCCT CAGCCTCCCA	4496
	AAGTGCTGGA ATTACAGATG ATTTCTTATT TCAGAAATCT	4536
	GCCAACTATA AAAGAGCAAT CTCTTGATAC TGTCTTGTCT	4576
10	GCTTCTCTTG CTTTCTCAAC CTCTTCTCAT TCTCTTTTTT	4616
	CTTTATATAT AATATATGTA TTTATATACA TACACTATAT	4656
15	ATACATTTTT GTATGCATTA TGCACTCATG TACGCAAAAA	4696
	GTTCTGAAAG TTGTCCTACA ATTTACTGTT TTATTTGCAT	4736
	ATTCAGACTT TGGCATTCCT GGACTCTATT CTTTAAAGAT	4776
20	TTGTTTTTCTG TGTGTTTCAA CATTCCTTTG TGGATTTAGG	4816
	ACAGTACACC TGCCAATTTT TATTCCAGGG ATGGATTCCA	4856
25	TTGTCACATT TCTGCAGTCA TTTCTCAGGG AGGGTTTTAA	4896
	GGTGGTGTTT TCCAAATGAC TTTTAAAAAA TATTTGAGAA	4936
	TGAGTATGCT TTTGTTATTG TTTGCTTATT TGTGCTTGG	4976
30	AAAACCTCTG TTCATGACCT TTATAGTTAT GCCCAGTTAT	5016
	TTTAAGGGTT CTCACTTCAG CAGTGGTTGT AAGCATCTGC	5056
35	CTCCCCTGAC ATTTTAAAAA TCCAGTTATC CCCACCAATT	5096
	AAGGAAGAAA AAGCTCAGTA AAGTTTATGC TGGTGCAACC	5136
	AACGTGCTTT AAACCTATCC TCAGGAATGG GAGGCAAAGC	5176
40	GACAGGTGGA AACATCTCAG GCTTAAAATA AGATATATCA	5216
	GAGTTCAAAT TCTGGATGGA TTGTTTACTT AAGGTGACCA	5256
45	TAAAATGTAT TATCTAAATT AGGACAATTA TAAGGGTAAA	5296
	AGAGTTCACT ACAAATAATT GGGAGGTATT AATTAACCTA	5336
50	TGATGCCAAT ATACTGGAAC TATTAAGTAG TCGTGTGATA	5376

55

	ATGAGCAAGT TGTCAAAATT TCTGTGTTTC AGCTTTCTTA	5416
5	GCTCTAAAAA AGGGCTACTA AATCCTGGAA ACATTTTCAT	5456
	AAGCATTAGT GATAATGTAT GACTCAGCAC CTGGCATAGA	5496
	GAAGCTCAAT AAATGGCAAC TGCTAATCAT CAAAACCACA	5536
10	GATCGGTAGT AGCTGTAGCT GCCACCTCTT CCTTAGGAGT	5576
	ATCACCTCTT TTAGGCAAAG CCATTACTTT GCCTGGTTTT	5616
15	CAAAAATGTG TTTATCATCT CAGTCTAGTG AAGAATAAAG	5656
	TGACATAATT GAGATCACCT AAGACATCAG CCAAATATCA	5696
	PvuII	
20	* GCTGGGCCTA TTGCTGACAT CATTCCAACA CCTTCTCGAT	5736
	CAATAA AAT TCT CTC TCA CCA AGT GGC TTT GTC CCC CTC	5775
	Asn Ser Leu Ser Pro Ser Gly Phe Val Pro Leu	
25	GTT AGA TTG CTC CCT TTC TAT AAA GTG GTT TGG CCA TAT	5814
	Val Arg Leu Leu Pro Phe Tyr Lys Val Val Trp Pro Tyr	
30	TTA CGC CAG TAT TGT ATA ATT TTA GAT TTA TCA AGC TGT	5853
	Leu Arg Gln Tyr Cys Ile Ile Leu Asp Leu Ser Ser Cys	
	GCA GGG AGA TGT GGG GAA GGG TAT TCT AGA GAT GCC ACC	5892
	Ala Gly Arg Cys Gly Glu Gly Tyr Ser Arg Asp Ala Thr	
35	TGC AAC TGT GAT TAT AAC TGT CAA CAC TAC ATG GAG TGC	5931
	Cys Asn Cys Asp Tyr Asn Cys Gln His Tyr Met Glu Cys	
	TGC CCT GAT TTC AAG AGA GTC TGC ACT GCG GGTAAGTCCT	5971
40	Cys Pro Asp Phe Lys Arg Val Cys Thr Ala	
	GAGAGCGGGT GTCTCCTCTG TCAAGCAACA CTGCGAGTCT	6011
	GTGAGTCCCC CCTTGCACCC TCGTGCAATG CTGTGAGACT	6051
45	GAGCCTCCCC TTGCACCCAC TTGCAGTGCT GTTTTCCCAC	6091
	TGTATCACTT CCTTTGCTTA AGTGACTTTC CTTCACTTGC	6131
50	ATAAATGTTG CAGCGCATTG TACCTAAGGA TGGTGTCAAA	6171
	GCTTGTGATT ATTGGGGAGG GTAAGGGGAG CCTGGAAGTC	6211

55

	TGTCACCTTGC	ACAAGGTTAT	TGTCTAGTTC	CAAGTAAATG	6251
5	TCTTATTTTA	AATAGAGAAG	ATACTGTGAG	TCAGGAGTCA	6291
	TTAAGTAAAT	AGTAAGGAGT	AATTATTTAC	AATCAGAGAT	6331
	GTCCTCGACT	AGTAATGAGG	ACTTTCAATA	GTTTCATCAA	6371
10	AAGAATACAT	GAGCATGGAA	TAATCTTGTC	TACCATTGGA	6411
	CTTTCTGATG	TTAACACGTT	TGTAAATTAC	CTGGAAAAGA	6451
15	CACTTTAAAG	ATAAAGTGCT	AGTGTTTAAT	GATGAATTTT	6491
	AGTTAAAATC	TGAAGGTCTG	AATTTTAAAT	TATGAGTGGT	6531
	GTTTCACTTG	AATATATTTT	GGGCCATTTT	ATTATTAACA	6571
20	CAGAAACATT	TCCTATTCAA	TTTAAGAAAA	TTGTAATTAA	6611
	AGTTATTTGA	AAGGTGTATT	TGCTTTAGAA	TTTGAATGTG	6651
25	AAGTTGAGGA	ATATATTGTG	CAATTTATGT	ATTGGATAGG	6691
	TAATTATTTT	AAATAAGCCT	TGGTAAGTCC	CTGTAACTC	6731
	TAACAAAGGC	TTTTTAAATT	TTCATTTTTT	AAATGAGTAG	6771
30	TCCTCCCTAA	CATAGTCCAC	ACTGTAAGAT	TAGGCTGAAA	6811
	GCTTTCAACT	ATACACCTTC	CCCAGTTTGT	CATAATAAGG	6851
35	GCCTCTGCAT	TACAAATGAT	TTTAAGCCTC	AAAAATGACC	6891
	CATTTACGTG	GAATATATAT	ACATATATAT	TTATATGGAA	6931
40		BglII			
		*			
	TATATTGCTG	AGATCTGTAC	ATTCAATTAC	TGTGAATCTA	6971
	TTACAAAGCA	GTGTGTGAAG	AGGAGAAGGA	TGAAGAGATT	7011
45	TCATATGAAG	GCTATCTCAC	TATCTAGACA	TTTCCCGATT	7051
	TTTCTTTGTC	CATACATGTA	AATAACTCGG	GCAGCATCAG	7091
50	GATGTCTCTT	GGAGTCTGGA	AGGGCAAGAG	GAGTTGCCCT	7131
	CAGTCACCAT	ATTTCTTTTT	TGACTTGGGC	TGTCTCCATC	7171

55

TGGGATACCA TCTAATTTTT CCTGGATGAT GTACTCCAAA 7211
 5 TTTCAAATAA AAGACTTAGA AATGAACTTT TGGAAACCTA 7251
 GTCAAGTCTA AGGTGGGAAA TGGCTGTCAA ATACGTGGGC 7291
 CTGGCTTCAC AATGAATAAT CTGTAACTTC TTGTTTTGCT 7331
 10 CTGGGTA GAG CTT TCC TGT AAA GGC CGC TGC TTT GAG TCC 7371
 Glu Leu Ser Cys Lys Gly Arg Cys Phe Glu Ser
 TTC GAG AGA GGG AGG GAG TGT GAC TGC GAC GCC CAA TGT 7410
 15 Phe Glu Arg Gly Arg Glu Cys Asp Cys Asp Ala Gln Cys
 AAG AAG TAT GAC AAG TGC TGT CCC GAT TAT GAG AGT TTC 7449
 Tyr Lys Lys Asp Lys Cys Cys Pro Asp Tyr Glu Ser Phe
 20 TGT GCA GAA GGTAAGCATC ACAGTACCAA CCAATGCTTC 7488
 Cys Ala Glu
 BglII
 *
 25 TCAGTACAGC CAGATCT 7505

30

35

40

45

50

55

TABLE IA

Reading frame A - (nucleotides #1 - 7503)

5 nl
 Asp Leu Phe Thr Leu Lys Asp Phe Leu Tyr Ser Phe Arg Pro
 Trp Ala Glu Met Tyr Thr Leu Leu Val Tyr Val Asp Arg Gln
 10 Ile Cys Asn Leu End Thr Ile Ile Ser Asn Phe Gln Glu Lys
 Lys Thr Tyr Trp Leu Gly Leu Phe Lys Phe Gln Phe Leu Ile
 Leu Ser Ala Val Ala Asp Val Gly Arg End His Ile End Pro
 Glu Leu Lys Gly Lys End Pro Asp Arg Lys Arg Ser Ser Ser
 Gln Lys Arg Gly Ile Ile End Arg End Ala Asp Ser Leu Glu
 15 Ile Val Phe Thr Tyr Leu Val Leu End Asn Lys His Cys Arg
 Lys Lys Ile Asp Ala Tyr Asn Ser Ser Leu Lys Lys Val Thr
 Ile Ser Gln Thr Tyr Val Tyr Met End Ser Lys Gly Val Lys
 Phe Ser Ile Met Ile Thr Asn End Val Met Val Leu Ile Lys
 Arg Glu Leu Asp Val Val Glu Thr End Arg Asn Leu Lys Cys
 Lys Asp Met Gly Ser Asn Leu Gly Ser Gly Leu End Ser Leu
 20 End Ile Leu Gly Trp Val Thr His Ser Ser Phe Val Gly Leu
 His Leu Leu Ile Cys Lys Cys Gly End Leu Tyr Leu Thr Leu
 Arg His Phe Val Arg Ile Lys Cys Ser Thr Tyr Gln Met Ser
 Ser Arg Gln Gln Asn Pro Ala Tyr Asn His Val Val His End
 Thr Cys Phe Leu Phe Phe Leu Phe Phe Phe Phe End Asp
 25 End Val Leu Leu Leu Leu Pro Lys Leu Val Cys Gly Gly Val
 Ile Ser Ala His Cys Asn Leu Cys Leu Ser Gly Ser Ser Asp
 Ser Pro Ala Leu Ala Ser Arg Val Thr Gly Ile Thr Gly Ala
 Cys His His Thr Arg Leu Ile Phe Cys Ile Phe Ser Arg Glu
 Gly Val Ser Pro Cys Trp Ser Gly Trp Ser Gln Thr Pro Asp
 30 Leu Arg End Ser Ala His Leu Ser Leu Pro Lys Cys Trp Asp
 Cys Arg Gly Glu Pro Leu His Pro Ala End Ala Cys Phe Lys
 His Ile Thr Leu Val Lys Gln Leu Gly Cys Met Leu Ala Ile
 Asp End His Ile Gly Val Gly Cys Pro End Gly Asn His Ser
 Leu Leu Gly Val End Asp Arg End Arg Asp Ile Leu Val Lys
 35 Arg Ser Thr Ile Val Tyr Tyr Asn Ser Arg Cys Asp Trp Asn
 Leu Ile Trp Lys Gly Phe Arg Glu Val Leu Gly His Glu Arg
 Ser Leu Met Leu Val Gly Lys Ile Leu His Arg Tyr Ile Pro
 Tyr Glu Ser Gly Val Lys Pro Trp Ala Glu Leu Thr End Leu
 Cys Asn Gly Asp Gly Ser His Phe His Cys Gly Lys Glu Thr
 40 Lys Leu Val Ile End Lys Thr Phe Ser Lys His Ala Tyr Leu
 Lys Gly Arg Ile End Gly Lys Tyr Gln Tyr Ile Phe End Ala
 Asp Phe End Asn Asn Lys Gly Ser End Lys Ile Gln Cys Gln
 End Ser Thr Cys Arg End Leu Arg Ala Cys Phe Leu Cys Ser
 Glu Lys Met Lys Leu Leu Lys Lys Cys Ile Phe Asp Ile Met
 45 Val End Val Tyr Lys Met Tyr Ala Thr Ser End His End Val
 Asn Phe Lys Ala Thr Phe Ala Arg Phe Asn Trp Lys Ala Glu
 Glu Ile Asn Asp Thr Ser Met End Ile Phe End Ser Cys Cys
 Phe Asn Ile Leu Val Lys Val Cys Gln Ser Tyr Leu Pro Ser
 Gln Asp Asn Leu Phe Thr Ser Leu Arg Asn End Lys Cys Ala
 50 Leu Ala Asn Thr Ser Lys Phe Ile Leu Trp End Ser Phe Cys
 Leu Met Glu Val Phe Ser Arg Cys Cys Leu Ser Leu Val Val
 Trp Ala Leu Phe Gln End Met Arg Lys Val End Phe His Arg

55

	Leu	Ser	Leu	Leu	Ile	Leu	Ser	Asn	Leu	Trp	Thr	Asn	Trp	Phe
	Ile	Pro	Leu	Tyr	Leu	His	End	Tyr	Ile	Arg	Leu	Pro	Gly	Arg
5	Gly	Ile	Asn	Ser	Ser	Ile	Leu	Tyr	Gly	End	Val	Ser	Lys	Leu
	Thr	Cys	Phe	Ile	Glu	Leu	Asn	Val	Phe	Ser	Leu	Glu	Val	Phe
	Glu	Lys	Lys	Cys	Val	Lys	Ile	Val	Val	Gly	Asn	Phe	End	Lys
	Gly	Leu	Cys	Leu	Cys	Asn	Thr	Pro	Val	Gly	Ser	Leu	End	Ile
	Tyr	Ile	Glu	Asn	End	Glu	Lys	Val	Gly	Thr	Val	Thr	Pro	End
10	Cys	Leu	Leu	Asn	Tyr	Ile	Ser	Val	Thr	Gln	Val	Ile	Ile	Phe
	Leu	Lys	Ile	Lys	Phe	End	Pro	Cys	Thr	Val	Lys	Tyr	Met	Pro
	Tyr	Phe	Gly	End	Ser	Phe	Lys	Ile	Ser	Leu	Phe	Tyr	Ile	Gly
	Asp	Phe	Leu	Arg	Ile	Lys	Thr	Val	Asn	Asn	Val	Ala	Glu	Gly
	Asp	Leu	Ile	Cys	Ser	Ala	End	Asp	End	Val	End	Trp	Pro	Asn
15	Ile	Phe	Tyr	Lys	Ile	Lys	Ala	Asp	Leu	Lys	Ser	Cys	Phe	Lys
	Asn	Gly	Met	Val	Leu	Leu	Leu	Leu	His	Phe	End	Ser	Tyr	End
	Asn	Val	Tyr	Asn	Ala	Ser	Leu	Ser	End	Trp	End	Asp	Glu	Arg
	Ala	Val	Phe	End	Tyr	Phe	Tyr	Phe	Ile	Phe	Ser	Lys	Gly	Thr
	Tyr	Gly	Thr	End	Lys	Gln	Arg	Trp	His	Gly	Lys	His	Phe	Pro
20	Phe	Thr	Cys	Cys	Cys	Cys	Cys	Leu	Phe	Ser	End	Phe	Ser	Lys
	Phe	His	Leu	Lys	Val	Ala	End	Pro	Ser	Asn	Ile	Leu	Leu	Phe
	Asn	Asn	Tyr	Cys	End	Ser	Phe	Ser	Leu	Asp	Phe	Tyr	Asn	Asn
	Gly	Asn	Ile	Phe	Leu	Lys	Ile	Tyr	Ile	Cys	Leu	Ser	Leu	Ile
	Tyr	Cys	Thr	Thr	Leu	Lys	End	Tyr	Ile	Thr	Cys	Ser	Leu	Phe
25	Val	Ser	Ala	Leu	Glu	Phe	Ser	Ser	Phe	Ser	Pro	Phe	Glu	Ser
	Phe	Glu	Leu	Val	End	Arg	Ile	Ile	Leu	Val	Leu	Val	Ile	Ser
	Met	Val	Leu	Ser	Lys	Val	End	Lys	Val	Gln	Thr	Val	His	His
	Lys	Ile	Met	Leu	Glu	Leu	Tyr	Ser	Ile	Ile	Lys	Ile	Thr	Lys
	Lys	Glu	Gln	Tyr	Leu	Asp	Asn	Met	End	Phe	Leu	Pro	Ile	Cys
30	End	Thr	Ile	Cys	Phe	Leu	Lys	Ser	Asn	Lys	Leu	Leu	Ile	Asn
	Asn	Leu	Phe	Trp	His	His	Leu	Asp	Ile	Asp	Thr	Ile	Leu	End
	End	Thr	Ile	Val	Phe	Gln	Met	Ala	Phe	Tyr	End	Ser	End	Lys
	Cys	Ser	Leu	Phe	Leu	Thr	Arg	Thr	Ile	Tyr	Pro	Gly	Met	Asp
	Ile	Tyr	Arg	Cys	Asp	Cys	His	Phe	Val	Leu	End	His	End	Thr
35	Phe	His	Ser	Ser	Pro	Lys	Thr	Gly	Pro	Gly	End	Ile	Val	Leu
	Val	Thr	His	Phe	End	Arg	Leu	Leu	Lys	Tyr	Ile	Gln	End	Asn
	His	Ile	Ser	Lys	Arg	Cys	End	His	Ala	Ile	End	Gly	Asp	End
	End	End	Gln	Cys	Gln	Ala	Leu	Asp	Phe	Ser	Phe	Leu	Phe	Phe
	Ser	Phe	Cys	Ser	Pro	Leu	Ile	Ser	Lys	Ile	End	Glu	Asn	Lys
40	Asn	Gln	Asp	Ser	End	Gln	Leu	End	End	Asn	Tyr	Val	Ser	Lys
	Val	Asn	Ile	Ser	Ser	Lys	Ile	Arg	Asp	Ser	Tyr	Glu	Thr	End
	Thr	End	Val	Val	Thr	Val	Cys	Ile	Gly	Ile	Ser	Cys	Arg	Ile
	Val	End	Ile	Arg	Lys	Asp	Arg	Lys	Thr	His	Glu	Ser	Lys	Cys
	Glu	Tyr	Ser	Ile	Arg	Met	Met	Thr	Thr	Pro	Val	Ser	Pro	Asp
45	Pro	Ser	Ala	Phe	Ser	End	His	Tyr	Asp	Lys	Val	His	Asn	Thr
	Gly	Gly	Thr	Phe	End	Thr	Ala	Phe	Ser	Ser	Val	Ile	Gln	Leu
	Val	Val	His	End	Asn	Ser	Val	Thr	Phe	Ile	Leu	Ile	Val	Ile
	Lys	End	Ile	Lys	Leu	Lys	Leu	Met	Gly	Leu	Gly	Val	Val	Ala
	His	Thr	Cys	Asn	Pro	Ser	Ile	Leu	Gly	Gly	Leu	Gly	Arg	Gln
50	Ile	Thr	End	Gly	Leu	Glu	Phe	Lys	Ser	Ser	Leu	Gly	Asn	Met
	Ala	Lys	Ser	Cys	Leu	Tyr	Thr	Lys	Tyr	Lys	Lys	Ile	Ser	Gln
	Val	Trp	End	Cys	Val	Pro	Val	Ile	Pro	Ala	Ile	Arg	Arg	Leu

55

Arg Gln Glu Asp Ile Leu Asn Pro Gly Gly Gly Gly Cys Ile
 Lys Leu Ser Leu His His Gly Thr Pro Ala Trp Ala Thr Glu
 End Asp Ser Val Ser Lys Ile Asn Lys End Ile Asn Lys Ile
 Asn Gly Tyr Asn End End Gln End Lys End Asp Arg His Leu
 Ile Tyr Glu Ser Lys Leu Phe Leu End Gly End Phe Pro Leu
 Phe Phe Phe Ser Phe Glu Thr Gly Ser Cys Phe Val Ser Gln
 Ser Gly Val Gln Trp His Asn Leu Gly Ser Leu Gln Pro Pro
 Pro Thr Gly Phe Lys Arg Phe Ser Cys Leu Ser Leu Pro Ser
 Ser Trp Asp Tyr Arg Pro Leu Pro Pro Cys Pro Val Asn Phe
 Cys Ile Phe Ser Lys Asp Arg Phe End Pro Cys Trp Pro Gly
 Trp Ser His Leu Thr Ser Gly Asp Pro Pro Ala Ser Ala Ser
 Gln Ser Ala Gly Ile Thr Asp Asp Phe Leu Phe Gln Lys Ser
 Ala Asn Tyr Lys Arg Ala Ile Ser End Tyr Cys Leu Val Cys
 Phe Ser Cys Phe Leu Asn Leu Phe Ser Phe Ser Phe Phe Phe
 Ile Tyr Asn Ile Cys Ile Tyr Ile His Thr Leu Tyr Ile His
 Phe Cys Met His Tyr Ala Leu Met Tyr Ala Lys Ser Ser Glu
 Ser Cys Pro Thr Ile Tyr Cys Phe Ile Cys Ile Phe Arg Leu
 Trp His Ser Trp Thr Leu Phe Phe End Asp Leu Phe Ser Val
 Cys Phe Asn Ile Pro Leu Trp Ile End Asp Ser Thr Pro Ala
 Asn Phe Tyr Ser Arg Asp Gly Phe His Cys His Ile Ser Ala

n4897

*

Val Ile Ser Gln Gly Gly Phe End Gly Gly Val Phe Gln Met
Thr Phe Lys Lys Tyr Leu Arg Met Ser Met Leu Leu Leu Leu
Phe Ala Tyr Leu Leu Pro Gly Lys Pro Leu Phe Met Thr Phe
Ile Val Met Pro Ser Tyr Phe Lys Gly Ser His Phe Ser Ser

n5073

*

Gly Cys Lys His Leu Pro Pro Leu Thr Phe End Lys Ser Ser
 Tyr Pro His Gln Leu Arg Lys Lys Lys Leu Ser Lys Val Tyr
 Ala Gly Ala Thr Asn Val Leu End Thr Tyr Pro Gln Glu Trp
 Glu Ala Lys Arg Gln Val Glu Thr Ser Gln Ala End Asn Lys
 Ile Tyr Gln Ser Ser Asn Ser Gly Trp Ile Val Tyr Leu Arg
 End Pro End Asn Val Leu Ser Lys Leu Gly Gln Leu End Gly
 End Lys Ser Ser Leu Gln Ile Ile Gly Arg Tyr End Leu Thr
 Tyr Asp Ala Asn Ile Leu Glu Leu Leu Thr Ser Arg Val Ile
 Met Ser Lys Leu Ser Lys Phe Leu Cys Phe Ser Phe Leu Ser
 Ser Lys Lys Gly Leu Leu Asn Pro Gly Asn Ile Phe Ile Ser
 Ile Ser Asp Asn Val End Leu Ser Thr Trp His Arg Glu Ala
 Gln End Met Ala Thr Ala Asn His Gln Asn His Arg Ser Val
 Val Ala Val Ala Ala Thr Ser Ser Leu Gly Val Ser Pro Ser
 Leu Gly Lys Ala Ile Thr Leu Pro Gly Phe Gln Lys Cys Val
 Tyr His Leu Ser Leu Val Lys Asn Lys Val Thr End Leu Arg
 Ser Pro Lys Thr Ser Ala Lys Tyr Gln Leu Gly Leu Leu Leu

n5742

*

Thr Ser Phe Gln His Leu Leu Asp Gln End Asn Ser Leu Ser
 Pro Ser Gly Phe Val Pro Leu Val Arg Leu Leu Pro Phe Tyr
 Lys Val Val Trp Pro Tyr Leu Arg Gln Tyr Cys Ile Ile Leu
 Asp Leu Ser Ser Cys Ala Gly Arg Cys Gly Glu Gly Tyr Ser
 Arg Asp Ala Thr Cys Asn Cys Asp Tyr Asn Cys Gln His Tyr

n5961

*

Met Glu Cys Cys Pro Asp Phe Lys Arg Val Cys Thr Ala Gly
 Lys Ser End Glu Arg Val Ser Pro Leu Ser Ser Asn Thr Ala
 Ser Leu End Val Pro Pro Cys Thr Leu Val Gln Cys Cys Glu
 Thr Glu Pro Pro Leu Ala Pro Thr Cys Ser Ala Val Phe Pro
 Leu Tyr His Phe Leu Cys Leu Ser Asp Phe Pro Ser Leu Ala
 End Met Leu Gln Arg Ile Val Pro Lys Asp Gly Val Lys Ala
 Cys Asp Tyr Trp Gly Gly End Gly Glu Pro Gly Ser Leu Ser
 Leu Ala Gln Gly Tyr Cys Leu Val Pro Ser Lys Cys Leu Ile
 Leu Asn Arg Glu Asp Thr Val Ser Gln Glu Ser Leu Ser Lys
 End End Gly Val Ile Ile Tyr Asn Gln Arg Cys Pro Arg Leu
 Val Met Arg Thr Phe Asn Ser Phe Ile Lys Arg Ile His Glu
 His Gly Ile Ile Leu Ser Thr Ile Gly Leu Ser Asp Val Asn
 Thr Phe Val Asn Tyr Leu Glu Lys Thr Leu End Arg End Ser
 Ala Ser Val End End End Ile Ser Val Lys Ile End Arg Ser
 Glu Phe End Ile Met Ser Gly Val Ser Leu Glu Tyr Ile Leu
 Gly His Phe Ile Ile Asn Thr Glu Thr Phe Pro Ile Gln Phe
 Lys Lys Ile Val Ile Lys Val Ile End Lys Val Tyr Leu Leu
 End Asn Leu Asn Val Lys Leu Arg Asn Ile Leu Cys Asn Leu
 Cys Ile Gly End Val Ile Ile Ser Asn Lys Pro Trp End Val
 Pro Val Asn Ser Asn Lys Gly Phe Leu Asn Phe His Phe Leu
 Asn Glu End Ser Ser Leu Thr End Ser Thr Leu End Asp End
 Ala Glu Ser Phe Gln Leu Tyr Thr Phe Pro Ser Leu Ser End
 End Gly Pro Leu His Tyr Lys End Phe Leu Ser Leu Lys Asn
 Asp Pro Phe Thr Trp Asn Ile Tyr Thr Tyr Ile Phe Ile Trp
 Asn Ile Leu Leu Arg Ser Val His Ser Ile Thr Val Asn Leu
 Leu Gln Ser Ser Val End Arg Gly Glu Gly End Arg Asp Phe
 Ile End Arg Lys Ser His Tyr Leu Asp Ile Ser Arg Phe Phe
 Phe Val His Thr Cys Lys End Leu Gly Gln His Gln Asp Val
 Ser Trp Ser Leu Glu Gly Gln Glu Glu Leu Pro Ser Val Thr
 Ile Phe Leu Phe End Leu Gly Leu Ser Pro Ser Gly Ile Pro
 Ser Asn Phe Ser Trp Met Met Tyr Ser Lys Phe Gln Ile Lys
 Asp Leu Glu Met Asn Phe Trp Lys Pro Ser Gln Val End Gly
 Gly Lys Trp Leu Ser Asn Thr Trp Ala Trp Leu His Asn Glu
 End Ser Val Thr Ser Cys Phe Ala Leu Gly Arg Ala Phe Leu
 End Arg Pro Leu Leu End Val Leu Arg Glu Arg Glu Gly Val
 End Leu Arg Arg Pro Met End Glu Val End Gln Val Leu Ser
 Arg Leu End Glu Phe Leu Cys Arg Arg End Ala Ser Gln Tyr
 Gln Pro Met Leu Leu Ser Thr Ala Arg ?

TABLE IB

Reading frame B - (nucleotides #2 - 7504)

n2

*

	Ile	Phe	Leu	Leu	End	Arg	Thr	Phe	Ser	Thr	Pro	Leu	Asp	His
10	Gly	Gln	Lys	Cys	Thr	His	Tyr	Trp	Ser	Thr	End	Thr	Asp	Lys
	Phe	Val	Ile	Ser	Glu	Leu	End	Phe	Gln	Ile	Ser	Arg	Arg	Arg
	Lys	His	Ile	Gly	Ser	Gly	Cys	Ser	Ser	Ser	Asn	Ser	End	Ser
	Tyr	Gln	Leu	Trp	Leu	Met	Trp	Glu	Asp	Asn	Ile	Tyr	Asn	Gln
	Ser	End	Lys	Glu	Asn	Ser	Pro	Ile	Glu	Arg	Glu	Val	Val	Pro
15	Arg	Lys	Gly	Gly	Leu	Phe	Arg	Asp	Glu	Gln	Ile	Ala	Ser	Lys
	Leu	Ser	Leu	His	Ile	End	Ser	Cys	Glu	Ile	Ser	Thr	Val	Glu
	Arg	Lys	End	Met	Leu	Ile	Ile	Leu	Val	Leu	Lys	Lys	Ser	Gln
	Ser	His	Lys	Pro	Met	Tyr	Thr	Cys	Arg	Val	Lys	Glu	End	Asn
	Leu	Val	End	End	Leu	Gln	Ile	Lys	End	Trp	Phe	End	Ser	Lys
20	Gly	Ser	Trp	Met	End	Trp	Lys	His	Asp	Gly	Thr	End	Asn	Ala
	Lys	Thr	Trp	Val	Gln	Ile	Leu	Val	Leu	Ala	Ser	Ser	Leu	Tyr
	Arg	Ser	Leu	Asp	Gly	Ser	His	Ile	Leu	His	Leu	Trp	Val	Phe
	Ile	Phe	Leu	Phe	Val	Asn	Val	Asp	Asp	Tyr	Ile	End	Leu	End
	Asp	Ile	Leu	End	Glu	Leu	Asn	Ala	Val	Pro	Thr	Lys	End	Val
25	Val	Asp	Ser	Arg	Ile	Gln	Pro	Thr	Thr	Met	Trp	Phe	Thr	Glu
	His	Val	Phe	Phe	Phe	Phe	Phe	Ser	Phe	Phe	Phe	Phe	Glu	Thr
	Glu	Phe	Cys	Ser	Cys	Cys	Pro	Ser	Trp	Cys	Ala	Val	Ala	End
	Ser	Gln	Leu	Thr	Ala	Thr	Ser	Ala	Ser	Gln	Val	Gln	Val	Ile
	Leu	Leu	Pro	End	Pro	Pro	Glu	End	Leu	Glu	Leu	Gln	Ala	Pro
30	Ala	Thr	Thr	Pro	Gly	End	Phe	Phe	Val	Phe	Leu	Val	Glu	Arg
	Glu	Phe	His	His	Val	Gly	Gln	Ala	Gly	Leu	Lys	Leu	Leu	Thr
	Ser	Gly	Asp	Leu	Pro	Thr	Ser	Ala	Ser	Gln	Ser	Ala	Gly	Ile
	Ala	Gly	Val	Ser	His	Cys	Thr	Arg	Arg	Glu	His	Val	Leu	Asn
	Ile	End	His	Trp	End	Asn	Asn	Trp	Ala	Val	Cys	Trp	Gln	End
35	Thr	Asn	Ile	End	Val	Leu	Val	Val	Leu	Arg	Glu	Ile	Ile	Val
	Cys	Trp	Gly	Cys	Glu	Thr	Asp	Glu	Glu	Ile	Phe	End	End	Asn
	Val	Val	Gln	Leu	Phe	Thr	Ile	Ile	Leu	Asp	Val	Ile	Gly	Ile
	End	Ser	Gly	Lys	Val	Ser	Gly	Arg	Cys	End	Asp	Met	Ser	Glu
	Ala	Leu	Cys	Trp	Trp	Val	Arg	Phe	Tyr	Thr	Asp	Thr	Phe	Arg
40	Met	Lys	Val	Gly	End	Ser	Leu	Gly	Leu	Ser	End	Leu	Ser	Ser
	Ala	Met	Val	Met	Val	His	Ile	Phe	Ile	Val	Gly	Arg	Lys	Gln
	Asn	End	End	Phe	Lys	Arg	Pro	Phe	Pro	Asn	Met	Leu	Ile	Leu
	Lys	Glu	Gly	Ser	Lys	Glu	Asn	Ile	Asn	Thr	Phe	Phe	Lys	Gln
	Ile	Phe	Lys	Ile	Ile	Lys	Glu	Ala	Arg	Lys	Ser	Ser	Val	Asn
45	Asn	Leu	Pro	Val	Gly	Asn	End	Glu	His	Val	Phe	Cys	Ala	Gln
	Arg	Lys	End	Asn	Tyr	Ser	Arg	Asn	Val	Tyr	Leu	Ile	End	Trp
	Tyr	Lys	Tyr	Ile	Lys	Cys	Met	Pro	Gln	Ala	Asn	Thr	Arg	End
	Ile	Ser	Lys	Leu	Pro	Leu	Leu	Asp	Leu	Thr	Gly	Lys	Gln	Lys
	Lys	End	Met	Ile	Pro	Val	Cys	Glu	Tyr	Phe	Lys	Val	Ala	Ala
50	Ser	Ile	Tyr	End	Ser	Lys	Ser	Ala	Arg	Ala	Ile	Tyr	His	Leu
	Arg	Thr	Ile	Cys	Ser	Leu	His	Leu	Glu	Ile	Arg	Asn	Val	Leu
	Trp	Pro	Ile	His	Leu	Asn	Leu	Phe	Cys	Gly	Ser	His	Phe	Ala

55

End Trp Lys Cys Phe Pro Asp Val Ala End Val End Ser Ser
 Gly Pro Phe Ser Asn Lys End Gly Lys Phe Asp Phe Ile Gly
 5 Cys His Cys End Phe Cys Leu Thr Phe Gly Leu Ile Gly Ser
 Ser His Cys Ile Cys Thr Asp Ile End Asp Ser Gln Asp Val
 Gly End Thr His Leu Ser Phe Thr Gly Lys Cys Gln Ser End
 Leu Ala Ser End Asn End Met Cys Phe His End Arg Cys Leu
 Lys Lys Asn Val End Arg End End Leu Glu Ile Phe Glu Lys
 10 Asp Tyr Val Tyr Ala Ile His Leu Leu Glu Ala Phe Glu Phe
 Ile Leu Arg Ile Lys Lys Lys Leu Glu Gln End Leu His Asp
 Ala Tyr End Ile Thr Phe Leu End His Arg Leu Leu Phe Ser
 Leu Arg End Asn Phe Asn Leu Ala Leu Leu Ser Thr Cys His
 Thr Leu Ala Arg Val Leu Arg End Val Tyr Ser Ile Leu Glu
 15 Ile Ser End Glu Leu Lys Leu Ser Ile Thr Trp Leu Arg Val
 Thr End Phe Val Leu His Lys Ile Lys Ser Asn Gly Gln Ile
 Phe Ser Ile Lys Leu Lys Leu Ile End Ser Leu Val Ser Arg
 Met Gly Trp Phe Tyr Tyr Cys Tyr Ile Phe Lys Val Thr Glu
 Met Cys Ile Met Gln Ala End Val Ser Gly Glu Met Lys Glu
 20 Leu Phe Ser Asp Thr Phe Ile Leu Phe Ser Ala Arg Val Pro
 Thr Val Pro Glu Asn Asn Asp Gly Met Glu Asn Thr Ser His
 Leu Pro Val Val Ala Ala Val Cys Phe Arg Asp Ser Ala Ser
 Phe Ile Ser Arg End Leu Asn His Arg Thr Tyr Phe Tyr Leu
 Thr Thr Ile Ala Asn His Ser Val Leu Ile Phe Ile Thr Thr
 25 Glu Ile Tyr Phe End Lys Phe Ile Phe Ala End Val End Tyr
 Thr Val Gln Pro End Asn Asn Ile End His Val Ala Cys Leu
 End Val Leu Trp Asn Ser His Leu Phe Leu Leu Leu Arg Val
 Leu Ser Leu Tyr Lys Glu End Phe Trp Tyr Leu Leu Phe Leu
 Trp Phe Ser Arg Lys Phe Lys Lys Cys Arg Gln Phe Ile Ile
 30 Lys Leu Cys Trp Ser Tyr Ile Val Ser End Lys End Leu Lys
 Lys Asn Asn Ile End Ile Thr Cys Asn Phe Cys Leu Tyr Val
 Glu Gln Phe Val Phe End Asn Gln Ile Asn Tyr End End Ile
 Thr Tyr Phe Gly Ile Ile End Thr End Ile Pro Tyr Tyr Asn
 Lys Leu End Cys Phe Lys Trp His Phe Ile Ser Leu Glu Asn
 35 Val Leu Phe Phe Leu Gln Glu Leu Phe Ile Gln Val Trp Thr
 Tyr Thr Asp Val Thr Val Ile Leu Cys Tyr Asn Ile Lys Leu
 Phe Thr Val Leu Pro Lys Gln Val Leu Asp Glu End Phe Leu
 Leu Leu Ile Phe Arg Asp Cys End Ser Thr Phe Ser Glu Ile
 Thr Leu Val Lys Asp Ala Ser Met Pro Ser Arg Gly Ile Asn
 40 Ser Asn Asn Ala Lys Leu Trp Ile Phe Leu Phe Phe Ser Phe
 Pro Phe Val Pro Leu End End Ala Lys Ser Glu Arg Ile Lys
 Ile Lys Ile His Asp Ser Tyr Asp Glu Ile Met Phe Leu Lys
 End Thr Phe Pro Val Lys Tyr Glu Ile Leu Met Lys Pro Glu
 His Glu Trp End Leu Ser Ala End Ala End Val Ala Glu Leu
 45 Phe Arg End Glu Lys Thr Gly Lys His Met Lys Ala Asn Val
 Asn Ile Gln End Glu End End Leu Leu Gln Tyr Leu Gln Thr
 Leu Arg Leu Ser Arg Asn Thr Met Thr Arg Phe Thr Thr Leu
 Gly Ala Leu Ser Lys Leu Pro Phe Pro Leu End Tyr Asn Trp
 Leu Phe Thr Lys Thr Val Leu Leu Ser Phe End Leu End Leu
 50 Asn Lys Ser Asn End Asn End Trp Gly Trp Val Trp Trp Leu
 Thr Pro Val Ile Pro Ala Phe Trp Glu Ala End Gly Asp Arg

55

Ser Pro Glu Val Trp Ser Ser Arg Ala Ala Trp Ala Thr Trp
 Gln Asn Leu Val Ser Ile Gln Asn Thr Lys Lys Leu Ala Arg
 Cys Gly Ser Val Cys Leu End Ser Gln Leu Phe Gly Gly End
 Gly Lys Arg Ile Ser End Ile Gln Glu Val Glu Val Ala Leu
 Ser End Ala Cys Ile Met Ala Leu Gln Pro Gly Gln Gln Ser
 Glu Thr Leu Ser Gln Lys End Ile Asn Lys End Ile Lys Leu
 Met Val Thr Ile Asn Ser Asn Lys Ser Arg Thr Asp Thr End

n4202

*

Ser Met Lys Val Ser Phe Ser Cys Lys Asp Asp Phe Leu Phe
 Phe Phe Phe Leu Leu Arg Gln Gly Leu Ala Leu Ser Pro Ser
 Leu Glu Cys Ser Gly Thr Ile Leu Ala His Cys Asn Leu His
 Leu Arg Gly Ser Ser Asp Ser Pro Val Ser Ala Ser Arg Val
 Ala Gly Ile Thr Gly Pro Cys His His Ala Arg Leu Ile Phe
 Val Phe Leu Val Lys Thr Gly Phe Asn His Val Gly Gln Ala

n4465 n4466

* *

Gly Leu Thr End Pro Gln Val Ile Arg Pro Pro Gln Pro Pro
 Lys Val Leu Glu Leu Gln Met Ile Ser Tyr Phe Arg Asn Leu
 Pro Thr Ile Lys Glu Gln Ser Leu Asp Thr Val Leu Ser Ala
 Ser Leu Ala Phe Ser Thr Ser Ser His Ser Leu Phe Ser Leu
 Tyr Ile Ile Tyr Val Phe Ile Tyr Ile His Tyr Ile Tyr Ile
 Phe Val Cys Ile Met His Ser Cys Thr Gln Lys Val Leu Lys
 Val Val Leu Gln Phe Thr Val Leu Phe Ala Tyr Ser Asp Phe
 Gly Ile Pro Gly Leu Tyr Ser Phe Lys Ile Cys Phe Gln Cys
 Val Ser Thr Phe Leu Cys Gly Phe Arg Thr Val His Leu Pro
 Ile Ser Ile Pro Gly Met Asp Ser Ile Val Thr Phe Leu Gln

n4915

*

Ser Phe Leu Arg Glu Gly Phe Lys Val Val Phe Ser Lys End
 Leu Leu Lys Asn Ile End Glu End Val Cys Phe Cys Tyr Cys
 Leu Leu Ile Cys Cys Leu Glu Asn Leu Cys Ser End Pro Leu
 End Leu Cys Pro Val Ile Leu Arg Val Leu Thr Ser Ala Val
 Val Val Ser Ile Cys Leu Pro End His Phe Lys Asn Pro Val
 Ile Pro Thr Asn End Gly Arg Lys Ser Ser Val Lys Phe Met
 Leu Val Gln Pro Thr Cys Phe Lys Leu Ile Leu Arg Asn Gly
 Arg Gln Ser Asp Arg Trp Lys His Leu Arg Leu Lys Ile Arg
 Tyr Ile Arg Val Gln Ile Leu Asp Gly Leu Phe Thr End Gly
 Asp His Lys Met Tyr Tyr Leu Asn End Asp Asn Tyr Lys Gly
 Lys Arg Val His Tyr Lys End Leu Gly Gly Ile Asn End Pro
 Met Met Pro Ile Tyr Trp Asn Tyr End Leu Val Val End End
 End Ala Ser Cys Gln Asn Phe Cys Val Ser Ala Phe Leu Ala
 Leu Lys Lys Gly Tyr End Ile Leu Glu Thr Phe Ser End Ala
 Leu Val Ile Met Tyr Asp Ser Ala Pro Gly Ile Glu Lys Leu
 Asn Lys Trp Gln Leu Leu Ile Ile Lys Thr Thr Asp Arg End
 End Leu End Leu Pro Pro Leu Pro End Glu Tyr His Pro Leu
 End Ala Lys Pro Leu Leu Cys Leu Val Phe Lys Asn Val Phe
 Ile Ile Ser Val End End Arg Ile Lys End His Asn End Asp

His Leu Arg His Gln Pro Asn Ile Ser Trp Ala Tyr Cys End
 His His Ser Asn Thr Phe Ser Ile Asn Lys* (nucleotide
 #5743)...

[Reading frame for nucleotides #5742 - #5961
 (Exon I) has been determined as reading frame A]

(nucleotide 5963) Val* Ser Pro Glu Ser Gly Cys Leu Leu
 Cys Gln Ala Thr Leu Arg Val Cys Glu Ser Pro Leu Ala Pro
 Ser Cys Asn Ala Val Arg Leu Ser Leu Pro Leu His Pro Leu
 Ala Val Leu Phe Ser His Cys Ile Thr Ser Phe Ala End Val
 Thr Phe Leu His Leu His Lys Cys Cys Ser Ala Leu Tyr Leu
 Arg Met Val Ser Lys Leu Val Ile Ile Gly Glu Gly Lys Gly
 Ser Leu Glu Val Cys His Leu His Lys Val Ile Val End Phe
 Gln Val Asn Val Leu Phe End Ile Glu Lys Ile Leu End Val
 Arg Ser His End Val Asn Ser Lys Glu End Leu Phe Thr Ile
 Arg Asp Val Leu Asp End End End Gly Leu Ser Ile Val Ser
 Ser Lys Glu Tyr Met Ser Met Glu End Ser Cys Leu Pro Leu
 Asp Phe Leu Met Leu Thr Arg Leu End Ile Thr Trp Lys Arg
 His Phe Lys Asp Lys Val Leu Val Phe Asn Asp Glu Phe Gln
 Leu Lys Ser Glu Gly Leu Asn Phe Lys Leu End Val Val Phe
 His Leu Asn Ile Phe Trp Ala Ile Leu Leu Leu Thr Gln Lys
 His Phe Leu Phe Asn Leu Arg Lys Leu End Leu Lys Leu Phe
 Glu Arg Cys Ile Cys Phe Arg Ile End Met End Ser End Gly
 Ile Tyr Cys Ala Ile Tyr Val Leu Asp Arg End Leu Phe Gln
 Ile Ser Leu Gly Lys Ser Leu Leu Thr Leu Thr Lys Ala Phe
 End Ile Phe Ile Phe End Met Ser Ser Pro Pro End His Ser
 Pro His Cys Lys Ile Arg Leu Lys Ala Phe Asn Tyr Thr Pro
 Ser Pro Val Cys His Asn Lys Gly Leu Cys Ile Thr Asn Asp
 Phe End Ala Ser Lys Met Thr His Leu Arg Gly Ile Tyr Ile
 His Ile Tyr Leu Tyr Gly Ile Tyr Cys End Asp Leu Tyr Ile
 Gln Leu Leu End Ile Tyr Tyr Lys Ala Val Cys Glu Glu Glu
 Lys Asp Glu Glu Ile Ser Tyr Glu Gly Tyr Leu Thr Ile End
 Thr Phe Pro Asp Phe Ser Leu Ser Ile His Val Asn Asn Ser
 Gly Ser Ile Arg Met Ser Leu Gly Val Trp Lys Gly Lys Arg
 Ser Cys Pro Gln Ser Pro Tyr Phe Phe Phe Asp Leu Gly Cys
 Leu His Leu Gly Tyr His Leu Ile Phe Pro Gly End Cys Thr
 Pro Asn Phe Lys End Lys Thr End Lys End Thr Phe Gly Asn
 Leu Val Lys Ser Lys Val Gly Asn Gly Cys Gln Ile Arg Gly
 Pro Gly Phe Thr Met Asn Asn Leu End Leu Leu Val Leu Leu

n7340

*

Trp Val Glu Leu Ser Cys Lys Gly Arg Cys Phe Glu Ser Phe
Glu Arg Gly Arg Glu Cys Asp Cys Asp Ala Gln Cys Lys Lys

n7459

*

Tyr Asp Lys Cys Cys Pro Asp Tyr Glu Ser Phe Cys Ala Glu
Gly Lys His His Ser Thr Asn Gln Cys Phe Ser Val Gln Pro
 Asp ?

TABLE IC

Reading frame C - (nucleotides #3 - 7505)

n3

*

5	Ser	Phe	Tyr	Ser	Glu	Gly	Leu	Ser	Leu	Leu	Leu	End	Thr	Met
	Gly	Arg	Asn	Val	His	Ile	Ile	Gly	Leu	Arg	Arg	Gln	Thr	Asn
10	Leu	End	Ser	Leu	Asn	Tyr	Asn	Phe	Lys	Phe	Pro	Gly	Glu	Glu
	Asn	Ile	Leu	Ala	Arg	Val	Val	Gln	Val	Pro	Ile	Pro	Asn	Pro
	Ile	Ser	Cys	Gly	End	Cys	Gly	Lys	Ile	Thr	Tyr	Ile	Thr	Arg
	Ala	Lys	Arg	Lys	Ile	Ala	Arg	End	Lys	Glu	Lys	End	Phe	Pro
	Glu	Lys	Gly	Asp	Tyr	Leu	Glu	Met	Ser	Arg	End	Pro	Arg	Asn
15	Cys	Leu	Tyr	Ile	Phe	Ser	Pro	Val	Lys	End	Ala	Leu	End	Lys
	Glu	Asn	Arg	Cys	Leu	End	Phe	End	Ser	End	Lys	Ser	His	Asn
	Leu	Thr	Asn	Leu	Cys	Ile	His	Val	Glu	End	Arg	Ser	Lys	Ile
	End	Tyr	Asn	Asp	Tyr	Lys	Leu	Ser	Asp	Gly	Phe	Asp	Gln	Lys
	Gly	Ala	Gly	Cys	Ser	Gly	Asn	Met	Thr	Glu	Leu	Lys	Met	Gln
20	Arg	His	Gly	Phe	Lys	Ser	Trp	Phe	Trp	Pro	Leu	Val	Phe	Ile
	Asp	Pro	Trp	Met	Gly	His	Thr	Phe	Phe	Ile	Cys	Gly	Ser	Ser
	Ser	Ser	Tyr	Leu	End	Met	Trp	Met	Ile	Ile	Phe	Asn	Phe	Lys
	Thr	Phe	Cys	Glu	Asn	End	Met	Gln	Tyr	Leu	Pro	Asn	Glu	End
	End	Thr	Ala	Glu	Ser	Ser	Leu	Gln	Pro	Cys	Gly	Ser	Leu	Asn
25	Met	Phe	Ser	Phe	Phe	Ser	Phe	Leu	Phe	Phe	Phe	Leu	Arg	Leu
	Ser	Phe	Ala	Leu	Val	Ala	Gln	Ala	Gly	Val	Arg	Trp	Arg	Asp
	Leu	Ser	Ser	Leu	Gln	Pro	Leu	Pro	Leu	Arg	Phe	Lys	End	Phe
	Ser	Cys	Leu	Ser	Leu	Pro	Ser	Asn	Trp	Asn	Tyr	Arg	Arg	Leu
	Pro	Pro	His	Pro	Ala	Asn	Phe	Leu	Tyr	Phe	End	End	Arg	Gly
30	Ser	Phe	Thr	Met	Leu	Val	Arg	Leu	Val	Ser	Asn	Ser	End	Pro
	Gln	Val	Ile	Cys	Pro	Pro	Gln	Pro	Pro	Lys	Val	Leu	Gly	Leu
	Gln	Gly	End	Ala	Thr	Ala	Pro	Gly	Val	Ser	Met	Phe	End	Thr
	Tyr	Asn	Ile	Gly	Glu	Thr	Thr	Gly	Leu	Tyr	Val	Gly	Asn	Arg
	Leu	Thr	Tyr	Arg	Cys	Trp	Leu	Ser	Leu	Gly	Lys	Ser	End	Ser
35	Val	Gly	Gly	Val	Arg	Gln	Met	Lys	Arg	Tyr	Phe	Ser	Lys	Thr
	End	Tyr	Asn	Cys	Leu	Leu	End	Phe	End	Met	End	Leu	Glu	Phe
	Asp	Leu	Glu	Arg	Phe	Gln	Gly	Gly	Ala	Arg	Thr	End	Ala	Lys
	Pro	Tyr	Val	Gly	Gly	End	Asp	Phe	Thr	Pro	Ile	His	Ser	Val
	End	Lys	Trp	Gly	Lys	Ala	Leu	Gly	End	Ala	Asp	Leu	Ala	Leu
40	Gln	Trp	End	Trp	Phe	Thr	Phe	Ser	Leu	Trp	Glu	Gly	Asn	Lys
	Ile	Ser	Asn	Leu	Lys	Asp	Leu	Phe	Gln	Thr	Cys	Leu	Ser	End
	Arg	Lys	Asp	Leu	Arg	Lys	Ile	Ser	Ile	His	Phe	Leu	Ser	Arg
	Phe	Leu	Lys	End	End	Arg	Lys	Leu	Glu	Asn	Pro	Val	Ser	Ile
	Ile	Tyr	Leu	End	Val	Thr	Glu	Ser	Met	Phe	Ser	Val	Leu	Arg
45	Glu	Asn	Glu	Thr	Thr	Gln	Glu	Met	Tyr	Ile	End	Tyr	Asn	Gly
	Ile	Ser	Ile	End	Asn	Val	Cys	His	Lys	Leu	Thr	Leu	Gly	Lys
	Phe	Gln	Ser	Tyr	Leu	Cys	End	Ile	End	Leu	Glu	Ser	Arg	Arg
	Asn	Lys	End	Tyr	Gln	Tyr	Val	Asn	Ile	Leu	Lys	Leu	Leu	Leu
	Gln	Tyr	Thr	Ser	Gln	Ser	Leu	Pro	Glu	Leu	Phe	Thr	Ile	Ser
50	Gly	Gln	Phe	Val	His	Phe	Thr	End	Lys	Leu	Glu	Met	Cys	Ser
	Gly	Gln	Tyr	Ile	End	Ile	Tyr	Phe	Val	Val	Val	Ile	Leu	Leu
	Asn	Gly	Ser	Val	Phe	Gln	Met	Leu	Pro	Lys	Ser	Ser	Arg	Leu

55

Gly Pro Phe Pro Ile Asn Glu Glu Ser Leu Ile Ser End Val
 Val Thr Val Asp Phe Val End Pro Leu Asp End Leu Val His
 Pro Thr Val Phe Ala Leu Ile Tyr Lys Thr Pro Arg Thr Trp
 5 Asp Lys Leu Ile Tyr Pro Leu Arg Val Ser Val Lys Val Asn
 Leu Leu His Arg Ile Lys Cys Val Phe Ile Arg Gly Val End
 Lys Lys Met Cys Lys Asp Ser Ser Trp Lys Phe Leu Lys Arg
 Ile Met Phe Met Gln Tyr Thr Cys Trp Lys Pro Leu Asn Leu
 10 Tyr End Glu Leu Arg Lys Ser Trp Asn Ser Asn Ser Met Met
 Leu Ile Lys Leu His Phe Cys Asp Thr Gly Tyr Tyr Phe Pro
 End Asp Lys Ile Leu Thr Leu His Cys End Val His Ala Ile
 Leu Trp Leu Glu Phe End Asp Lys Ser Ile Leu Tyr Trp Arg
 Phe Leu Lys Asn End Asn Cys Gln End Arg Gly End Gly End
 15 Leu Asp Leu Phe Cys Ile Arg Leu Ser Leu Met Ala Lys Tyr
 Phe Leu End Asn End Ser End Phe Lys Val Leu Phe Gln Glu
 Trp Asp Gly Phe Ile Ile Val Thr Phe Leu Lys Leu Leu Lys

n2559

*

20 Cys Val End Cys Lys Pro Lys Leu Val Val Arg End Lys Ser
 Cys Phe Leu Ile Leu Leu Phe Tyr Phe Gln Gln Gly Tyr Leu
Arg Tyr Leu Lys Thr Thr Met Ala Trp Lys Thr Leu Pro Ile
Tyr Leu Leu Leu Leu Leu Ser Val Phe Val Ile Gln Gln Val

n2732

*

Ser Ser Gln Gly Ser Leu Thr Ile Glu His Thr Phe Ile End
 30 Gln Leu Leu Leu Ile Ile Gln Ser End Phe Leu End Gln Arg
 Lys Tyr Ile Ser Lys Asn Leu Tyr Leu Leu Glu Phe Asn Ile
 Leu Tyr Asn Leu Lys Ile Ile Tyr Asn Met End Pro Val Cys
 Lys Cys Phe Gly Ile Leu Ile Phe Phe Ser Phe End Glu Phe
 End Ala Cys Ile Lys Asn Asn Phe Gly Thr Cys Tyr Phe Tyr
 Gly Ser Leu Glu Ser Leu Lys Ser Ala Asp Ser Ser Ser End
 35 Asn Tyr Ala Gly Ala Ile End Tyr His Lys Asn Asn End Lys
 Arg Thr Ile Phe Arg End His Val Ile Phe Ala Tyr Met Leu
 Asn Asn Leu Phe Phe Lys Ile Lys End Ile Ile Asp Lys End
 Leu Ile Leu Ala Ser Phe Arg His Arg Tyr His Ile Ile Ile
 Asn Tyr Ser Val Ser Asn Gly Ile Leu Leu Val Leu Lys Met
 40 Phe Ser Phe Ser Tyr Lys Asn Tyr Leu Ser Arg Tyr Gly His
 Ile Gln Met End Leu Ser Phe Cys Val Ile Thr Leu Asn Phe
 Ser Gln Phe Ser Gln Asn Arg Ser Trp Met Asn Ser Ser Cys
 Tyr Ser Phe Leu Glu Thr Val Lys Val His Ser Val Lys Ser
 His End End Lys Met Leu Ala Cys His Leu Gly Gly Leu Ile
 45 Val Thr Met Pro Ser Phe Gly Phe Phe Phe Ser Phe Leu Phe
 Leu Leu Phe Pro Phe Asp Lys Gln Asn Leu Arg Glu End Lys
 Ser Arg Phe Met Thr Val Met Met Lys Leu Cys Phe End Ser
 Lys His Phe Gln End Asn Thr Arg Phe Leu End Asn Leu Asn
 Met Ser Gly Asn Cys Leu His Arg His Lys Leu Gln Asn Cys
 50 Leu Asp Lys Lys Arg Gln Glu Asn Thr End Lys Gln Met End
 Ile Phe Asn Lys Asn Asp Asp Tyr Ser Ser Ile Ser Arg Pro
 Phe Gly Phe Leu Val Thr Leu End Gln Gly Ser Gln His Trp

55

Gly His Phe Leu Asn Cys Leu Phe Leu Cys Asp Thr Ile Gly
 Cys Ser Leu Lys Gln Cys Tyr Phe His Phe Asn Cys Asp End
 Ile Asn Gln Ile Lys Ile Asn Gly Ala Gly Cys Gly Gly Ser
 5 His Leu End Ser Gln His Ser Gly Arg Pro Arg Glu Thr Asp
 His Leu Arg Ser Gly Val Gln Glu Gln Pro Gly Gln His Gly
 Lys Ile Leu Ser Leu Tyr Lys Ile Gln Lys Asn End Pro Gly
 Val Val Val Cys Ala Cys Asn Pro Ser Tyr Ser Glu Ala Glu
 10 Ala Arg Gly Tyr Leu Glu Ser Arg Arg Trp Arg Leu His End
 Ala Glu Leu Ala Ser Trp His Ser Ser Leu Gly Asn Arg Val
 Arg Leu Cys Leu Lys Asn Lys End Ile Asn Lys End Asn End
 Trp Leu Gln Leu Ile Ala Ile Lys Val Gly Gln Thr Pro Asn
 Leu End Lys End Ala Phe Pro Val Arg Met Ile Ser Ser Phe
 Phe Phe Phe Phe End Asp Arg Val Leu Leu Cys Leu Pro Val
 15 Trp Ser Ala Val Ala Gln Ser Trp Leu Thr Ala Thr Ser Thr
 Tyr Gly Val Gln Ala Ile Leu Leu Ser Gln Pro Pro Glu End
 Leu Gly Leu Gln Ala Pro Ala Thr Met Pro Gly End Phe Leu
 Tyr Phe End End Arg Gln Val Leu Thr Met Leu Ala Arg Leu
 Val Ser Pro Asp Leu Arg End Ser Ala Arg Leu Ser Leu Pro
 20 Lys Cys Trp Asn Tyr Arg End Phe Leu Ile Ser Glu Ile Cys
 Gln Leu End Lys Ser Asn Leu Leu Ile Leu Ser Cys Leu Leu
 Leu Leu Leu Ser Gln Pro Leu Leu Ile Leu Phe Phe Leu Tyr
 Ile End Tyr Met Tyr Leu Tyr Thr Tyr Thr Ile Tyr Thr Phe
 25 Leu Tyr Ala Leu Cys Thr His Val Arg Lys Lys Phe End Lys
 Leu Ser Tyr Asn Leu Leu Phe Tyr Leu His Ile Gln Thr Leu
 Ala Phe Leu Asp Ser Ile Leu Leu Arg Phe Val Phe Ser Val
 Phe Gln His Ser Phe Val Asp Leu Gly Gln Tyr Thr Cys Gln
 Phe Leu Phe Gln Gly Trp Ile Pro Leu Ser His Phe Cys Ser
 30 His Phe Ser Gly Arg Val Leu Arg Trp Cys Phe Pro Asn Asp
 Phe End Lys Ile Phe Glu Asn Glu Tyr Ala Phe Val Ile Val
 Cys Leu Phe Val Ala Trp Lys Thr Ser Val His Asp Leu Tyr
 Ser Tyr Ala Gln Leu Phe End Gly Phe Ser Leu Gln Gln Trp
 Leu End Ala Ser Ala Ser Pro Asp Ile Leu Lys Ile Gln Leu
 35 Ser Pro Pro Ile Lys Glu Glu Lys Ala Gln End Ser Leu Cys
 Trp Cys Asn Gln Arg Ala Leu Asn Leu Ser Ser Gly Met Gly
 Gly Lys Ala Thr Gly Gly Asn Ile Ser Gly Leu Lys End Asp
 Ile Ser Glu Phe Lys Phe Trp Met Asp Cys Leu Leu Lys Val
 Thr Ile Lys Cys Ile Ile End Ile Arg Thr Ile Ile Arg Val
 40 Lys Glu Phe Thr Thr Asn Asn Trp Glu Val Leu Ile Asn Leu
 End Cys Gln Tyr Thr Gly Thr Ile Asn End Ser Cys Asp Asn
 Glu Gln Val Val Lys Ile Ser Val Phe Gln Leu Ser End Leu
 End Lys Arg Ala Thr Lys Ser Trp Lys His Phe His Lys His
 End End End Cys Met Thr Gln His Leu Ala End Arg Ser Ser
 45 Ile Asn Gly Asn Cys End Ser Ser Lys Pro Gln Ile Gly Ser
 Ser Cys Ser Cys His Leu Phe Leu Arg Ser Ile Thr Leu Phe
 Arg Gln Ser His Tyr Phe Ala Trp Phe Ser Lys Met Cys Leu
 Ser Ser Gln Ser Ser Glu Glu End Ser Asp Ile Ile Glu Ile
 Thr End Asp Ile Ser Gln Ile Ser Ala Gly Pro Ile Ala Asp
 50 Ile Ile Pro Thr Pro Ser Arg Ser Ile* (nucleotide 5741)...

Coding sequence for Exon I occurs from
nucleotides #5742 - #5961....

5 (nucleotide 5961) Gly** End Val Leu Arg Ala Gly Val Ser
Ser Val Lys Gln His Cys Glu Ser Val Ser Pro Pro Leu His
Pro Arg Ala Met Leu End Asp End Ala Ser Pro Cys Thr His
Leu Gln Cys Cys Phe Pro Thr Val Ser Leu Pro Leu Leu Lys
10 End Leu Ser Phe Thr Cys Ile Asn Val Ala Ala His Cys Thr
End Gly Trp Cys Gln Ser Leu End Leu Leu Gly Arg Val Arg
Gly Ala Trp Lys Ser Val Thr Cys Thr Arg Leu Leu Ser Ser
Ser Lys End Met Ser Tyr Phe Lys End Arg Arg Tyr Cys Glu
Ser Gly Val Ile Lys End Ile Val Arg Ser Asn Tyr Leu Gln
15 Ser Glu Met Ser Ser Thr Ser Asn Glu Asp Phe Gln End Phe
His Gln Lys Asn Thr End Ala Trp Asn Asn Leu Val Tyr His
Trp Thr Phe End Cys End His Val Cys Lys Leu Pro Gly Lys
Asp Thr Leu Lys Ile Lys Cys End Cys Leu Met Met Asn Phe
Ser End Asn Leu Lys Val End Ile Leu Asn Tyr Glu Trp Cys
20 Phe Thr End Ile Tyr Phe Gly Pro Phe Tyr Tyr End His Arg
Asn Ile Ser Tyr Ser Ile End Glu Asn Cys Asn End Ser Tyr
Leu Lys Gly Val Phe Ala Leu Glu Phe Glu Cys Glu Val Glu
Glu Tyr Ile Val Gln Phe Met Tyr Trp Ile Gly Asn Tyr Phe
Lys End Ala Leu Val Ser Pro Cys End Leu End Gln Arg Leu
25 Phe Lys Phe Ser Phe Phe Lys End Val Val Leu Pro Asn Ile
Val His Thr Val Arg Leu Gly End Lys Leu Ser Thr Ile His
Leu Pro Gln Phe Val Ile Ile Arg Ala Ser Ala Leu Gln Met
Ile Phe Lys Pro Gln Lys End Pro Ile Tyr Val Glu Tyr Ile
Tyr Ile Tyr Ile Tyr Met Glu Tyr Ile Ala Glu Ile Cys Thr
30 Phe Asn Tyr Cys Glu Ser Ile Thr Lys Gln Cys Val Lys Arg
Arg Arg Met Lys Arg Phe His Met Lys Ala Ile Ser Leu Ser
Arg His Phe Pro Ile Phe Leu Cys Pro Tyr Met End Ile Thr
Arg Ala Ala Ser Gly Cys Leu Leu Glu Ser Gly Arg Ala Arg
Gly Val Ala Leu Ser His His Ile Ser Phe Leu Thr Trp Ala
35 Val Ser Ile Trp Asp Thr Ile End Phe Phe Leu Asp Asp Val
Leu Gln Ile Ser Asn Lys Arg Leu Arg Asn Glu Leu Leu Glu
Thr End Ser Ser Leu Arg Trp Glu Met Ala Val Lys Tyr Val
Gly Leu Ala Ser Gln End Ile Ile Cys Asn Phe Leu Phe Cys
Ser Gly End Ser Phe Pro Val Lys Ala Ala Ala Leu Ser Pro
40 Ser Arg Glu Gly Gly Ser Val Thr Ala Thr Pro Asn Val Arg
Ser Met Thr Ser Ala Val Pro Ile Met Arg Val Ser Val Gln
Lys Val Ser Ile Thr Val Pro Thr Asn Ala Ser Gln Tyr Ser
Gln Ile ?

45

50

55

TABLE II

Partial Clone Containing Exon III

5	TTTTATAGCA ATGCATCCTT AGCTTAAATG GAGTATTCGT	40
	GTTGAGCTGG AGCCTGCCTG CACTGGCTGT CACCAGCATC	80
10	TACTCTTGAA GCTAGATGCA TCTGTGCTTT TCACAGTTAG	120
	AGCTGCTGAT GAACATAAAC AAGATGTTAA CTGACTTGTC	160
15	TTACTTGGCC TCA GTG CAT AAT CCC ACA TCA CCA CCA TCT	200
	Val His Asn Pro Thr Ser Pro Pro Ser	
20	TCA AAG AAA GCA CCT CCA CCT TCA GGA GCA TCT CAA ACC	239
	Ser Lys Lys Ala Pro Pro Pro Ser Gly Ala Ser Gln Thr	
	ATC AAA TCA ACA ACC AAA CGT TCA CCC AAA CCA CCA AAC	278
	Ile Lys Ser Thr Thr Lys Arg Ser Pro Lys Pro Pro Asn	
25	AAG AAG AAG ACT AAG AAA GTT ATA GAA TCA GAG GAA ATA	317
	Lys Lys Lys Thr Lys Lys Val Ile Glu Ser Glu Glu Ile	
	ACA GAA GGT AGG AAG ATG ACA GAT ATA ATC AAA GGA GCT	356
	Thr Glu Gly Arg Lys Met Thr Asp Ile Ile Lys Gly Ala	
30	TTC TTA GAT GAA GTA ACT TGT AGG TGA CTGCTTA	390
	Phe Lys Asp Glu Val Thr Cys Arg	
	TCTAAGCCCA TTCTCAGAGA ACAGGGTAAT CTTAGGAATC	430
35	ATGAGCCTCA TTACACTCGA AGGTTTTAGA CTTTGCTTTT	470
	AAGTAAAGTT TAAGACAAAG TATAAACTCT CAGCTCTTTC	500
40	TGTATTTACC AAACCCAGCA TGAGTCTGGG TTAAAACAAA	540
	TCAGAAGGGA CAAATCTTAC TATAAAAAAC AAAAACCACC	580
	CCATGATTTT CTTTGTAGAA TAATTTGATT CTGTGTTTTG	620
45	GAGGATATGG GAAAGTTGAG AGATACTAGT AATACTGCTA	660
	GTATCTGTGA TAAGCCCAGG TGCCTTGCTT TTAAGTACA	700
50	GATTAAAAGG CAGTTGGTCA TATTACTAAT AAAAGCAAAA	740

55

EP 0 732 401 A2

	TCCAGATACT TGTAGACTAG TAAATAGCAC TTCTTGCTGT	780
5	GTTTAGACTG GTGGTTCTTT TTTGTTTAA ATCACAGTTG	820
	GTGTGATCC	829

10

15

20

25

30

35

40

45

50

55

TABLE III

Putative Partial cDNA of meg-CSF

(1)
 AAT TCT CTC TCA CCA AGT GGC TTT GTC CCC CTC GTT AGA TTG
 Asn Ser Leu Ser Pro Ser Gly Phe Val Pro Leu Val Arg Leu
 (20)
 CTC CCT TTC TAT AAA GTG GTT TGG CCA TAT TTA CGC CAG TAT
 Leu Pro Phe Tyr Lys Val Val Trp Pro Tyr Leu Arg Gln Tyr
 (30) (40)
 TGT ATA ATT TTA GAT TTA TCA AGC TGT GCA GGG AGA TGT GGG
 Cys Ile Ile Leu Asp Leu Ser Ser Cys Ala Gly Arg Cys Gly
 (50)
 GAA GGG TAT TCT AGA GAT GCC ACC TGC AAC TGT GAT TAT AAC
 Glu Gly Tyr Ser Arg Asp Ala Thr Cys Asn Cys Asp Tyr Asn
 (60) (70)
 TGT CAA CAC TAC ATG GAG TGC TGC CCT GAT TTC AAG AGA GTC
 Cys Gln His Tyr Met Glu Cys Cys Pro Asp Phe Lys Arg Val
 (80)
 TGC ACT GCG GAG CTT TCC TGT AAA GGC CGC TGC TTT GAG TCC
 Cys Thr Ala Glu Leu Ser Cys Lys Gly Arg Cys Phe Glu Ser
 (90)
 TTC GAG AGA GGG AGG GAG TGT GAC TGC GAC GCC CAA TGT AAG
 Phe Glu Arg Gly Arg Glu Cys Asp Cys Asp Ala Gln Cys Lys
 (100) (110)
 AAG TAT GAC AAG TGC TGT CCC GAT TAT GAG AGT TTC TGT GCA
 Lys Tyr Asp Lys Cys Cys Pro Asp Tyr Glu Ser Phe Cys Ala
 (120)
 GAA GTG CAT AAT CCC ACA TCA CCA CCA TCT TCA AAG AAA GCA
 Glu Val His Asn Pro Thr Ser Pro Pro Ser Ser Lys Lys Ala
 (130) (140)
 CCT CCA CCT TCA GGA GCA TCT CAA ACC ATC AAA TCA ACA ACC
 Pro Pro Pro Ser Gly Ala Ser Gln Thr Ile Lys Ser Thr Thr
 (150)
 AAA CGT TCA CCC AAA CCA CCA AAC AAG AAG AAG ACT AAG AAA
 Lys Arg Ser Pro Lys Pro Pro Asn Lys Lys Lys Thr Lys Lys

(160)

GTT ATA GAA TCA GAG GAA ATA ACA GAA GGT AGG AAG ATG ACA
 Val Ile Glu Ser Glu Glu Ile Thr Glu Gly Arg Lys Met Thr

(170)

GAT ATA ATC AAA GGA GCT TTC TTA GAT GAA GTA ACT TGT AGG
 Asp Ile Ile Lys Gly Ala Phe Lys Asp Glu Val Thr Cys Arg

(180)

Two genomic clones containing the meg-CSF sequences identified above have been deposited with the American Type Culture Collection, 12301 Parklawn Drive, Rockville, Maryland 20852, USA in accordance with the requirements of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the Purposes of Patent Procedure on August 3, 1990. An approximately 12 kb genomic fragment (referred to as Meg KpnI-SnaBI) spanning the 5' KpnI site to the 3' SnaBI site (see Fig. 1) in an *E. coli* plasmid was given the accession number ATCC 40857. The entire 18.2 kb NotI sequence of Fig. 1 (referred to as 18-5665) inserted into bacteriophage lambda DNA was deposited under the accession number ATCC 40856.

Preliminary experiments have shown that the KpnI/SnaBI fragment appears to produce an active protein in COS cells.

The nucleotide sequence of this meg-CSF sequence, specifically Exon I and Exon II, has been compared with the nucleotide sequences recorded in protein and DNA databanks. The amino terminus of vitronectin, the serum adhesion molecule, was observed to have the highest degree of sequence similarity. The amino terminal portion is also called Somatomedin B, a peptide found in the circulation. Other significant sequence similarities were found at the protein and DNA levels to the B domain of von Willebrand factor, made by endothelial cells and megakaryocytes and at the protein level in the extracellular domain of murine PC-1, a membrane glycoprotein dimer found on IgG secreting plasma cells. These peptides are functionally unrelated, with Somatomedin B and PC-1 having unknown functions.

To obtain the entire genomic and cDNA sequences the amino and carboxy termini of meg-CSF may be determined by various procedures. One procedure involves the preparation of a cDNA library from activated human peripheral blood leucocytes or other sources of meg-CSF RNA and extraction of the full length cDNA by hybridization, using the three known Exons as probes. A second method is expression cloning in COS cells. For example, different sections of the 18.2 kb genomic insert are subcloned into COS cells and different activities identified. If a protein is located, the RNA of that clone is isolated and cDNA prepared therefrom and expressed to obtain the protein. For example, this method involves subcloning either the full length human genomic clone (referred to as 18-5665), the KpnI/SnaBI clone, or different sections of the 18.2 kb genomic insert into an expression vector, transfecting into COS cells, preparing a cDNA library from meg-CSF transfected COS cells and screening by hybridization for meg-CSF cDNA. Alternatively, the entire sequence, including the N-terminal Met, may be identified by comparison with the murine homolog of meg-CSF. Also, the mRNA from a cell source of meg-CSF can be used to make a cDNA library which can be screened with the probes to identify the cDNAs encoding the meg-CSF polypeptide. Techniques to screen for cDNA sources include making primers from Exons I-III and employing PCR techniques to isolate and amplify cDNA transcripts.

The full-length human genomic clone or fragments thereof may also be employed as probes to isolate by cross-hybridization the murine genomic equivalent of meg-CSF. The murine genomic clone of meg-CSF or fragments thereof can be used to identify a mRNA source for meg-CSF which can be used to make a murine cDNA library. The murine cDNA can be used to identify the corresponding Exons in the human meg-CSF gene which can then be spliced together to create a full length human cDNA.

Once the entire cDNA is identified, it or any portion of it that encodes an active fragment of meg-CSF, can be introduced into any one of a variety of expression vectors to make an expression system for meg-CSF or one or more fragments thereof.

By such use of recombinant techniques, DNA sequences encoding the meg-CSF polypeptide are obtained which contain DNA sequences encoding one or more of the tryptic fragments or the partial sequence identified above. The present invention also encompasses these DNA sequences, free of association with DNA sequences encoding other proteins, and coding on expression for meg-CSF polypeptides. These DNA sequences include those sequences encoding all or a fragment of the above-identified peptide sequences or partial clone sequence and those sequences which hybridize under stringent hybridization conditions [see, T. Maniatis et al, Molecular Cloning (A Laboratory Manual), Cold Spring Harbor Laboratory (1982), pages 387 to 389] to the DNA sequences.

An example of one such stringent hybridization condition is hybridization in 4XSSC at 65°C, followed by a washing in 0.1XSSC at 65°C for an hour. Alternatively an exemplary stringent hybridization condition is in 50% formamide, 4XSSC at 42°C.

5 DNA sequences which hybridize to the sequences for meg-CSF under relaxed hybridization conditions and which code on expression for meg-CSF peptides having meg-CSF biological properties also encode novel meg-CSF polypeptides. Examples of such non-stringent hybridization conditions are 4XSSC at 50°C or hybridization with 30-40% formamide at 42°C. For example, a DNA sequence which shares regions of significant homology, e.g., sites of glycosylation or disulfide linkages, with the sequences of meg-CSF and encodes a protein having one or more meg-CSF biological properties clearly encodes a meg-CSF polypeptide even if such a DNA sequence would not stringently hybridize to the
10 meg-CSF sequences.

Allelic variations (naturally-occurring base changes in the species population which may or may not result in an amino acid change) of DNA sequences encoding the peptide sequences of meg-CSF are also included in the present invention, as well as analogs or derivatives thereof. Similarly, DNA sequences which code for meg-CSF polypeptides but which differ in codon sequence due to the degeneracies of the genetic code or variations in the DNA sequence of
15 meg-CSF which are caused by point mutations or by induced modifications to enhance the activity, half-life or production of the polypeptides encoded thereby are also encompassed in the invention.

meg-CSF polypeptides may also be produced by known conventional chemical synthesis. Methods for constructing the polypeptides of the present invention by synthetic means are known to those of skill in the art. The synthetically-constructed meg-CSF polypeptide sequences, by virtue of sharing primary, secondary, or tertiary structural and confor-
20 mational characteristics with meg-CSF polypeptides may possess meg-CSF biological properties in common therewith. Thus, they may be employed as biologically active or immunological substitutes for natural, purified meg-CSF polypeptides in therapeutic and immunological processes.

Modifications in the peptides or DNA sequences encoding meg-CSF can be made by one skilled in the art using known techniques. Modifications of interest in the meg-CSF sequences may include the replacement, insertion or dele-
25 tion of a selected amino acid residue in the coding sequences. Mutagenic techniques for such replacement, insertion or deletion are well known to one skilled in the art. [See, e.g., United States patent 4,518,584.]

Specific mutations of the sequences of the meg-CSF polypeptide may involve modifications of a glycosylation site. The absence of glycosylation or only partial glycosylation results from amino acid substitution or deletion at any asparagine-linked glycosylation recognition site or at any site of the molecule that is modified by addition of O-linked carbo-
30 hydrate. An asparagine-linked glycosylation recognition site comprises a tripeptide sequence which is specifically recognized by appropriate cellular glycosylation enzymes. These tripeptide sequences are either Asp-X-Thr or Asp-X-Ser, where X can be any amino acid. A variety of amino acid substitutions or deletions at one or both of the first or third amino acid positions of a glycosylation recognition site (and/or amino acid deletion at the second position) results in non-glycosylation at the modified tripeptide sequence. Expression of such altered nucleotide sequences produces var-
35 iants which are not glycosylated at that site.

Other analogs and derivatives of the sequence of meg-CSF which would be expected to retain meg-CSF activity in whole or in part may also be easily made by one of skill in the art given the disclosures herein. One such modification may be the attachment of polyethylene glycol (PEG) onto existing lysine residues in the meg-CSF sequence or the insertion of one or more lysine residues or other amino acid residues that can react with PEG or PEG derivatives into
40 the sequence by conventional techniques to enable the attachment of PEG moieties. Such modifications are believed to be encompassed by this invention.

The present invention also provides a method for producing meg-CSF polypeptides or active fragments thereof. One method of the present invention involves introducing the cDNA encoding a meg-CSF polypeptide into an expres-
45 sion vector to make an expression system for meg-CSF. A selected host cell is transformed with the vector and cultured. The method of this present invention therefore comprises culturing a suitable cell or cell line, which has been transformed with a DNA sequence coding on expression for a meg-CSF polypeptide under the control of known regulatory sequences. Regulatory sequences include promoter fragments, terminator fragments and other suitable sequences which direct the expression of the protein in an appropriate host cell. The expressed factor is then recovered, isolated and purified from the culture medium (or from the cell, if expressed intracellularly) by appropriate means known to one
50 of skill in the art.

Suitable cells or cell lines may be mammalian cells, such as Chinese hamster ovary cells (CHO) or 3T3 cells. The selection of suitable mammalian host cells and methods for transformation, culture, amplification, screening and prod-
55 uct production and purification are known in the art. See, e.g., Gething and Sambrook, *Nature*, 293:620-625 (1981), or alternatively, Kaufman et al, *Mol. Cell. Biol.*, 5(7):1750-1759 (1985) or Howley et al, U. S. Patent 4,419,446. Other suitable mammalian cell lines, are the monkey COS-1 cell line, and the CV-1 cell line. Further exemplary mammalian host cells include particularly primate cell lines and rodent cell lines, including transformed cell lines. Normal diploid cells, cell strains derived from *in vitro* culture of primary tissue, as well as primary explants, are also suitable. Candidate cells may be genotypically deficient in the selection gene, or may contain a dominantly acting selection gene. Other suitable

mammalian cell lines include but are not limited to, HeLa, mouse L-929 cells, 3T3 lines derived from Swiss, Balb-c or NIH mice, BHK or HaK hamster cell lines.

Similarly useful as host cells suitable for the present invention are bacterial cells. For example, the various strains of E. coli (e.g., HB101, MC1061 and strains used in the following examples) are well-known as host cells in the field of biotechnology. Various strains of B. subtilis, Pseudomonas, other bacilli and the like may also be employed in this method.

Many strains of yeast cells known to those skilled in the art are also available as host cells for expression of the polypeptides of the present invention. Additionally, where desired, insect cells may be utilized as host cells in the method of the present invention. See, e.g. Miller et al, Genetic Engineering, 8:277-298 (Plenum Press 1986) and references cited therein.

The present invention also provides recombinant molecules or vectors for use in the method of expression of novel meg-CSF polypeptides. These vectors contain the novel meg-CSF DNA sequences recited herein, and which alone or in combination with other sequences code for meg-CSF polypeptides of the invention or active fragments thereof. Alternatively, vectors incorporating modified sequences as described above are also embodiments of the present invention and useful in the production of meg-CSF polypeptides. The vector employed in the method also contains selected regulatory sequences in operative association with the DNA coding sequences of the invention and capable of directing the replication and expression thereof in selected host cells.

One desirable vector is pXM [Y. C. Yang et al, Cell, 47:3-10 (1986)]. Mammalian cell expression vectors described herein may be synthesized by techniques well known to those skilled in this art. The components of the vectors, e.g. replicons, selection genes, enhancers, promoters, and the like, may be obtained from natural sources or synthesized by known procedures. See, Kaufman et al, J. Mol. Biol., 159:511-521 (1982); and Kaufman, Proc. Natl. Acad. Sci. USA, 82:689-693 (1985). Alternatively, the vector DNA may include all or part of the bovine papilloma virus genome [Lusky et al, Cell, 36:391-401 (1984)] and be carried in cell lines such as C127 mouse cells as a stable episomal element. The transformation of these vectors into appropriate host cells can result in expression of the meg-CSF polypeptides.

Other appropriate expression vectors of which numerous types are known in the art for mammalian, insect, yeast, fungal and bacterial expression can also be used for this purpose.

Thus meg-CSF or active fragments thereof, purified to homogeneity from cell sources or produced recombinantly or synthetically, may be used in a pharmaceutical preparation or formulation to stimulate platelet recovery following chemotherapy or bone marrow transplantation, to treat thrombocytopenia, aplastic anemia and other platelet disorders. Therapeutic treatment of such platelet disorders or deficiencies with these meg-CSF polypeptide compositions may avoid undesirable side effects caused by treatment with presently available serum-derived factors or transfusions of human platelets. It may also be possible to employ one or more peptide fragments of meg-CSF, such as the peptides above-identified, in such pharmaceutical formulations.

The polypeptides of the present invention may also be employed, alone or in combination with other cytokines, hematopoietins, interleukins, growth factors or antibodies in the treatment of the above-identified conditions.

Therefore, as yet another aspect of the invention are therapeutic compositions for treating the conditions referred to above. Such compositions comprise a therapeutically effective amount of the meg-CSF protein or a therapeutically effective fragment thereof in admixture with a pharmaceutically acceptable carrier. This composition can be systematically administered parenterally. Alternatively, the composition may be administered intravenously. If desirable, the composition may be administered subcutaneously. When systematically administered, the therapeutic composition for use in this invention is in the form of a pyrogen-free, parenterally acceptable aqueous solution. The preparation of such pharmaceutically acceptable protein solutions, having due regard to pH, isotonicity, stability and the like, is within the skill of the art.

The dosage regimen involved in a method for treating the above-described conditions will be determined by the attending physician considering various factors which modify the action of drugs, e.g. the condition, body weight, sex and diet of the patient, the severity of any infection, time of administration and other clinical factors. Generally, the daily regimen should be in the range of 1-1000 micrograms of meg-CSF protein or fragment thereof or 50 to 5000 units (i.e., one unit being the minimum concentration of meg-CSF protein which yields the maximal number of colonies in the murine fibrin clot megakaryocyte colony formation assay) of protein per kilogram of body weight.

The therapeutic method, compositions and polypeptides of the present invention may also be employed, alone or in combination with other cytokines, hematopoietins, interleukins, growth factors or antibodies in the treatment of disease states characterized by other symptoms as well as platelet deficiencies. It is anticipated that this molecule, if it does not itself have TPO activity, will prove useful in treating some forms of thrombocytopenia in combination with general stimulators of hematopoiesis, such as IL-3, IL-6 or GM-CSF or with other megakaryocytic stimulatory factors or molecules with TPO-like activity. Additional exemplary cytokines or hematopoietins for such co-administration include TPO, G-CSF, CSF-1, GM-CSF, IL-1, IL-11 (described as IL-10 in co-owned copending U. S. patent application SN 07/441,100 incorporated herein by reference), IL-3, IL-4, M-CSF, IL-7 or erythropoietin. The dosage recited above would be adjusted to compensate for such additional components in the therapeutic composition. Progress of the treated patient can be monitored by conventional methods.

Other uses for these novel polypeptides are in the development of antibodies generated by standard methods for in vivo or in vitro diagnostic or therapeutic use. Such antibodies may include both monoclonal and polyclonal antibodies, as well as chimeric antibodies or "recombinant" antibodies generated by known techniques. Also provided by this invention are the cell lines generated by presenting meg-CSF or a fragment thereof as an antigen to a selected mammal, followed by fusing cells of the animal with certain cancer cells to create immortalized cell lines by known techniques. The methods employed to generate such cell lines and antibodies directed against all or portions of a human meg-CSF polypeptide of the present invention are also encompassed by this invention.

The antibodies of the present invention may be utilized for in vivo and in vitro diagnostic purposes, such as by associating the antibodies with detectable labels or label systems. Alternatively these antibodies may be employed for in vivo and in vitro therapeutic purposes, such as by association with certain toxic or therapeutic compounds or moieties known to those of skill in this art. These antibodies also have utility as research reagents.

The following examples illustratively describe the purification and characteristics of homogeneous human meg-CSF and other methods and products of the present invention. These examples are for illustration and do not limit the scope of the present invention.

Example 1 - Purification of meg-CSF from Urine

The following procedures are presently employed to obtain homogeneous meg-CSF protein from urine of human bone marrow transplant patients. Urine from patients with aplastic anemia or thrombocytopenia accompanying other disease states may also be used as the source of the factor employing this purification

STEP 1: Urine was collected from the bone marrow transplant patients between days 5 and 18 following transplant. One hundred liters of pooled urine were treated with protease inhibitors phenylmethyl- sulfonylfluoride (PMSF) and ethylenediaminetetraacetic acid (EDTA). This pooled urine was concentrated on an Amicon YM-10 filter (10,000 molecular weight cut-off) to remove excess pigments and reduce the volume. A cocktail of protease inhibitors (leupeptin, aprotinin, ethylene glycol-bis-tetraacetic acid (EGTA) and N-ethylmaleimide (NEM)) was added to the urine at this and the next three steps to minimize proteolysis. The pH of the urine concentrate was adjusted to 8.0 and diluted to a conductivity of 7mS/cm.

STEP 2: The retentate from this first step of the purification was then subjected to anion exchange column chromatography on a QAE Zetaprep [Cuno] at pH 8.0. The QAE flow-through was adjusted to a pH 4.5 with 1M acetic acid.

STEP 3: The flow-through from the second purification step was bound to a cation exchange chromatographic column, an SP-Zetaprep column [Cuno] at pH 4.5. Bound protein containing meg-CSF was eluted with 1M NaCl at a pH of 4.5. The eluate was pooled, protease inhibitors were added as above and the materials stored at -80°C until further chromatography was performed. The eluate was then dialyzed against Tris-buffered saline (TBS), with the addition of the protease inhibitors described in Step 1. This dialyzate was heated at 56°C for 30 minutes. Addition of the protease inhibitors, while not essential for recovery of protein, enabled greater amount of protein to be recovered from this step, undegraded by the proteases in the system. Pools from this step were also analyzed for the presence of megakaryocyte-specific growth factors. These pools were found to contain meg-CSF activity.

STEP 4: The resulting material was added to a lectin affinity chromatographic column, a Wheat Germ Sepharose column [Pharmacia] and eluted with 0.25 M N-acetyl glucosamine (N-acglcNH₂) in TBS. Urinary meg-CSF was found to bind to this column. The bound protein was eluted from this column by 20 mM sodium acetate, pH 4.5 in the presence of the protease inhibitors of Step 1, which were added for the reasons described in Step 3.

STEP 5: This dialysate was applied to a 10 ml S-Toyopearl FPLC cation exchange column and eluted using a linear gradient of 0 to 1M NaCl in 20mM sodium acetate at pH 4.5. The protein eluted from this step was tested for meg-CSF activity in the fibrin clot assay described below. The meg-CSF activity was observed to elute in two discrete peaks. The major activity eluted between 0.1M and 0.25M NaCl. A minor, but reproducible activity eluted between 0.3M and 0.5N NaCl. The two activities may be due to protein or carbohydrate modification of a single protein; however the data presented further herein refers to the major protein.

STEP 6: The eluate from this fifth purification step was then purified on a reverse phase HPLC (C4) column [Vydac; 1cmX25cm] which was eluted with a linear gradient of between 23-33% acetonitrile in 0.1% trifluoroacetic acid (TFA). This step removes an abundant 30Kd protein contaminant.

STEP 7: The HPLC step was repeated in a different solvent system, after the eluate of Step 6 was diluted with too parts acetic acid and pyridine. The purified material eluted between 6-15% n-propanol in pyridine and acetic acid on a C18 reverse phase HPLC column (0.46 X 25 cm). The material produced after this step, when assayed gave the specific activity of greater than 5 X 10⁷ dilution units reported in the murine assay. This optional step removes the bulk of urinary ribonuclease, a major contaminant, from the preparation.

STEP 8: The HPLC step was repeated once more on a C4 column (Vydac; 0.46 X 25 cm) using 0.15% HFBA in acetonitrile. The material eluted between 27-37% acetonitrile. The last HPLC step removed substantially all remaining ribonuclease and proteinaceous contaminants present after Step 7.

This purified meg-CSF material was then analyzed by SDS-PAGE, bioassayed and labelled with ^{125}I . Homogenous protein is obtained from this procedure, omitting step 7, having a specific activity ranging from about 5×10^7 to about 2×10^8 dilution units per mg protein in the murine megakaryocyte colony assay described below.

5 Example 2 - Sodium Dodecyl Sulfate-Polyacrylamide Gel Electrophoresis

The purified protein was analyzed by SDS-PAGE performed according to the method of Laemmli [Laemmli, U. K., *Nature*, 227:680-685 (1970)] on 12% acrylamide slab gels (0.75 mm thickness). After electrophoresis, the gels were either subjected to autoradiography to visualize ^{125}I -labelled meg-CSF, or silver stain, or cut into 0.5-1 cm slices and eluted in 0.5 ml TBS with 0.3% deionized BSA overnight at 4°C and assayed for meg-CSF activity. Apparent molecular weight was determined with protein standards: BRL prestained molecular weight markers, 14C molecular weight standards [NEN], or low molecular weight SDS-PAGE standards [Biorad].

A small aliquot of protein from Steps 6, 7 and 8 of Example 1 containing active meg-CSF was iodinated and subjected to SDS-PAGE. SDS-PAGE analysis (non-reducing conditions) of reverse phase purified meg-CSF from step 8 beginning with several fractions which eluted before the meg-CSF activity, continuing right through the active fractions and ending with fractions which eluted after the peak of meg-CSF activity, revealed the presence of one heterogenous protein band ranging in size between 28 and 38 kd. Elution of the protein from a parallel gel lane revealed that the bio-activity in the murine megakaryocyte colony formation assay correlated with the presence of the iodinated meg-CSF band in the gel.

Upon reduction, the majority of the protein has a molecular weight of between approximately 20-27kd. Based on this information meg-CSF may be a dimer. The protein does not appear to be digestable with N-glycanase under standard conditions.

25 Example 3 - Recovery of Protein

Starting with 50 liters of urine, the final pooled active fractions from the HPLC column contained approximately 25 micrograms of protein, estimated from the amino acid composition of purified meg-CSF. The specific activity of the 28-38 kd meg-CSF protein was estimated to be greater than approximately 5×10^7 dilution units/mg in the murine fibrin clot assay described below. One unit of activity is defined as the reciprocal of the maximal dilution which stimulates maximal colony formation. One megakaryocyte colony is defined as 3 or more cells.

30 Example 4 - meg-CSF Protein Composition

meg-CSF obtained from the eighth step of the purification of Example 1, omitting Step 7, was employed to obtain tryptic fragments for sequencing. Twenty-five micrograms of purified meg-CSF preparation were desalted over a reverse phase column. The main peak was then fully reduced and alkylated, due to the large number of cysteines present therein. This material was again eluted through a reverse phase column, and the protein peak subjected to SDS-PAGE. Using ^{125}I labelled material as a marker, protein running at approximately 20-27 kd was excised from the gel, fixed with methanol:acetic acid:water, and rinsed with water. The gel slice was macerated. Following neutralization with 0.1M NH_4HCO_3 solution (200 μl), the protein contained within the gel matrix was then digested with trypsin (2% w/w).

Sequencing provided the four peptide sequences:

- (a) Ser Arg Cys Phe Glu Ser Phe Glu Arg
- 45 (b) Arg Val Cys Thr Ala Glu Leu Ser Cys Lys Gly (Arg)
- (c) Lys Ala Pro Pro Pro (X) Gly Ala Ser Gln Thr Ile Lys
- (d) Lys Tyr Asp Lys Cys Cys Pro Asp Thr Glu Ser Phe Cys Ala Glu Val His Asn Pro

X represents an ambiguously identified amino acid, which is an S or T. () represents a tentatively identified amino acid.

50 All four of these tryptic peptides are found in the Exons and putative cDNA sequence of meg-CSF.

Example 5 - Genomic Cloning of Urinary meg-CSF.

Probes consisting of pools of oligonucleotides or unique oligonucleotides are designed from the tryptic sequences above according to the method of R. Lathe, *J. Mol. Biol.*, 183(1):1-12 (1985). The following oligonucleotide probes are synthesized on an automated DNA synthesizer, with N representing any of the four nucleotides A, T, C, or G; R representing the nucleotides A or G; Y representing the nucleotides C or T; and H representing the nucleotides A, T, or C:

(1) TGYTTYGARTCNTTYGA

(2) TGYTTYGARAGYTTYGA
 (3) GTNTGYACNGCNGARYT
 (4) AARGCNCCNCCNCCN
 (5) GCNAGYCARACNATHAA
 (6) GCNTCNCARACNATHAA
 (7) AARTAYGAYAARTGYTG
 (8) GCNGARGTNCAYAAAYCC
 (9)

AAGTATGACAAGTGCTGCCCTGATGAGTCCTTCTGTGCTGA
 GGTGCACAACCCC;

and
 (10)

AAGTATGACAAGTGCTGCCCTGATGAGAGCTTCTGTGCTGA
 GGTGCACAACCC

Because the genetic code is degenerate (more than one codon can code for the same amino acid) a mixture of oligonucleotides are synthesized that contain all possible nucleotide sequences encoding the amino acid sequence of the selected tryptic fragment or portion thereof. It may be possible in some cases to reduce the number of oligonucleotides in the probe mixture based on codon usage because some codons are rarely used in eukaryotic genes, and because of the relative infrequency of the dinucleotide CpG in eukaryotic coding sequences [see J. J. Toole et al, *Nature*, 312:342-347 (1984)]. The regions of the amino acid sequences used for probe design are chosen by avoiding highly degenerate codons where possible. The oligonucleotides are synthesized on an automated DNA synthesizer and the probes are then radioactively labelled with polynucleotide kinase and ³²P-ATP.

The degenerate oligonucleotide probes are then used to screen a human genomic library prepared from placenta [Stratagene Cloning Systems, La Jolla, CA] using established techniques [See K. Jacobs et al, *Nature*, 313:806-810 (1985)]. Recombinants from this library are plated and duplicate nitrocellulose and/or nylon replicas are made of the plates. Typically, the oligonucleotides are kinased with ³²P gamma ATP and hybridized to the filters at 48°C in 3M TMAC solution for 48-96 hours. The filters are then washed in 3M TMAC in 50 mM Tris at 50°C for 1 hour followed by two washes at room temperature for 30 minutes each in 2 x SSC [See K. Jacobs et al, *Nucleic Acids Res.*, 16:4637-4650 (1988)]. Duplicate positives are plaque purified.

An 18.2 kb genomic fragment (Fig. 1) was shown to hybridize to all four tryptic sequences. The first partial genomic subclone contained within the 18.2 kb insert to be isolated and sequenced was shown to contain one tryptic sequence (a) in a single Exon (Exon II) and two partial tryptic sequences (b and d) which are contained in Exon II and overlap with adjacent coding sequence. Exon III was identified with a probe made from tryptic (c). Exon I was identified with a probe made to the 5' end of tryptic (b).

The predicted cDNA and predicted amino acid sequences of these combined partial genomic clones are reported in Table III above, which encodes a portion of the meg-CSF protein. While the peptide encoded by this partial sequence may produce an active meg-CSF fragment, the sequence lacks the amino terminal Met as well as any defined 3' border. To obtain the remainder of the meg-CSF genomic sequence, the full length genomic DNA sequence may be expressed in COS cells and a cDNA library prepared from COS cell RNA, and the cDNA sequence cloned from that source. Alternatively, the remainder of the sequence may be deduced by sequence comparison with a cross-hybridizing murine genomic meg-CSF sequence. The sequence may also be obtained from peripheral blood lymphocytes or placenta, two potential sources of the mRNA.

Example 6 - Expression of Recombinant Human meg-CSF

To produce meg-CSF or an active fragment thereof, the cDNA encoding it is transferred into an appropriate expression vector, of which numerous types are known in the art for human, insect, yeast, fungal and bacterial expression, by standard molecular biology techniques. One such vector for mammalian cells is pXM [Y. C. Yang et al, *Cell*, 47:3-10 (1986)]. This vector contains the SV40 origin of replication and enhancer, the adenovirus major late promoter, a cDNA

copy of the adenovirus tripartite leader sequence, a small hybrid intervening sequence, an SV40 polyadenylation signal and the adenovirus VA I gene, in appropriate relationships to direct the high level expression of the desired cDNA in mammalian cells [See, e.g., Kaufman, *Proc. Natl. Acad. Sci. USA*, **82**:689-693 (1985)]. The pXM vector is linearized with the endonuclease enzyme XhoI and subsequently ligated in equimolar amount separately to the cDNA encoding meg-CSF modified by addition of synthetic oligonucleotides [Collaborative Research, Lexington, MA] that generate XhoI complementary ends to generate constructs for expression of meg-CSF.

Another vector which may be employed to express meg-CSF in CHO cells is pEMC2B1. This vector may be derived from pMT2pc which has been deposited with the American Type Culture Collection (ATCC), Rockville, MD (USA) under Accession Number ATCC 40348. The DNA is linearized by digestion of the plasmid with PstI. The DNA is then blunted using T₄ DNA polymerase. An oligonucleotide 5' TGCAGGCGAGCCTGAA TTCCTCGA 3' is then ligated into the DNA, recreating the PstI site at the 5' end and adding an EcoRI site and XhoI site before the ATG of the DHFR cDNA. This plasmid is called pMT21. pMT21 is cut with EcoRI and XhoI which cleaves the plasmid at two adjacent cloning sites. An EMCV fragment of 508 base pairs was cut from pMT₂ECAT₁ [S. K. Jong et al, *J. Virol.*, **63**:1651-1660 (1989)] with the restriction enzymes EcoRI and TaqαI. A pair of oligonucleotides 68 nucleotides in length were synthesized to duplicate the EMCV sequence up to the ATG. The ATG was changed to an ATT, and a C is added, creating a XhoI site at the 3' end. A TaqαI site is situated at the 5' end. The sequences of the oligonucleotides were:

5' CGAGGTTAAAAACGTCTAGGCCCCCGAACCACGGGGACGTGGTTCCTTT
GAAAAACACGATTGC

and its complementary strand.

Ligation of the pMT21 EcoRI-to-XhoI fragment to the EMCV EcoRI-to-TaqαI fragment and to the TaqαI/XhoI oligonucleotides produced the vector pEMC2B1. This vector contains the SV40 origin of replication and enhancer, the adenovirus major late promoter, a cDNA copy of the majority of the adenovirus tripartite leader sequence, a small hybrid intervening sequence, an SV40 polyadenylation signal and the adenovirus VA I gene, DHFR and β-lactamase markers and an EMC sequence, in appropriate relationships to direct the high level expression of the desired cDNA in mammalian cells. The EMC2B1 vector is linearized with the endonuclease enzyme EcoRI and subsequently ligated in equimolar amount separately to the cDNA encoding meg-CSF that was previously modified by addition of synthetic oligonucleotides that generate EcoRI complementary ends to generate constructs for expression.

The desired vector containing meg-CSF is then introduced into appropriate host cells by conventional genetic engineering techniques. The transformed cells are cultured and the expressed meg-CSF is recovered and purified from the culture medium using standard techniques.

A. Mammalian Cell Expression

To obtain expression of the meg-CSF polypeptide in mammalian host cells, the pXM vector containing the meg-CSF DNA sequence is transfected onto COS cells. The conditioned medium from the transfected COS cells contains meg-CSF biological activity as measured in the murine assays. Similarly the pEMC2B1 construct containing the cDNA for meg-CSF is transfected into CHO cells.

The mammalian cell expression vectors described herein may be synthesized by techniques well known to those skilled in this art. One skilled in the art can also construct other mammalian expression vectors comparable to the pXM vector by, e.g., inserting the DNA sequence of the meg-CSF from the plasmid with appropriate enzymes and employing well-known recombinant genetic engineering techniques and other known vectors, such as pJL3 and pJL4 [Gough et al., *EMBO J.*, **4**:645-653 (1985)] and pMT2 (starting with pMT2-VWF, ATCC #67122; see PCT application PCT/US87/00033).

Mammalian host cells other than COS cells may also be employed in meg-CSF expression. For example, preferably for stable integration of the vector DNA, and for subsequent amplification of the integrated vector DNA, both by conventional methods, CHO cells may be employed as a mammalian host cell of choice.

Once the vectors and host cells are selected and transformed, stable transformants are then screened for expression of the product by standard immunological, biological or enzymatic assays, such as those described below in Example 8. The presence of the DNA and mRNA encoding the meg-CSF polypeptides may be detected by standard procedures such as Southern and Northern blotting. Transient expression of the DNA encoding the polypeptides during the several days after introduction of the expression vector DNA into suitable host cells is measured without selection by activity or immunologic assay, e.g., the murine fibrin clot assay, of the proteins in the culture medium.

B. Bacterial Expression Systems

Similarly, one skilled in the art could manipulate the sequences encoding the meg-CSF polypeptide by eliminating any human regulatory sequences flanking the coding sequences and inserting bacterial regulatory sequences to create bacterial vectors for intracellular or extracellular expression of the meg-CSF polypeptide of the invention by bacterial cells. The DNA encoding the polypeptides may be further modified to contain different codons to optimize bacterial expression as is known in the art. Preferably the sequences encoding the mature meg-CSF are operatively linked in-frame to nucleotide sequences encoding a secretory leader polypeptide permitting bacterial expression, secretion and processing of the mature meg-CSF polypeptides, also by methods known in the art. The expression of meg-CSF in *E. coli* using such secretion systems is expected to result in the secretion of the active polypeptide. This approach has yielded active chimeric antibody fragments [See, e.g., Bitter et al, *Science*, 240:1041-1043 (1983)]. Alternatively, the meg-CSF may be expressed as a cytoplasmic protein in *E. coli*. In this case, the molecule would most likely have to be refolded after complete denaturation with guanidine hydrochloride, a process also known in the art. For procedures for isolation and refolding of intracellularly expressed proteins, see, for example, U. S. Patent 4,512,922.

The compounds expressed through either route in bacterial host cells may then be recovered, purified, and/or characterized with respect to physicochemical, biochemical and/or clinical parameters, all by known methods.

C. Insect or Yeast Cell Expression

Similar manipulations can be performed for the construction of an insect vector for expression of meg-CSF polypeptides in insect cells [See, e.g., procedures described in published European patent application 155,476].

Similarly yeast vectors are constructed employing yeast regulatory sequences to express cDNA encoding the precursor, in yeast cells to yield secreted extracellular active meg-CSF. Alternatively the polypeptide may be expressed intracellularly in yeast, the polypeptide isolated and refolded to yield active meg-CSF. [See, e.g., procedures described in published PCT application WO 86/00639 and European patent application EP 123,289.]

Example 7 - Construction of CHO Cell Lines Expressing High Levels of meg-CSF

One method for producing high levels of the meg-CSF protein of the invention from mammalian cells involves the construction of cells containing multiple copies of the cDNA encoding the meg-CSF.

The cDNA is co-transfected with an amplifiable marker, e.g., the DHFR gene for which cells containing increasing concentrations of methotrexate (MTX) according to the procedures of Kaufman and Sharp, *J. Mol. Biol.*, (1982) *supra*. This approach can be employed with a number of different cell types. Alternatively, the meg-CSF cDNA and drug resistance selection gene (e.g., DHFR) may be introduced into the same vector. A preferred vector for this approach is pEMC2B1.

For example, the pXM vector containing the meg-CSF gene in operative association with other plasmid sequences enabling expression thereof is introduced into DHFR-deficient CHO cells, DUKX-BII, along with a DHFR expression plasmid such as pAdD26SVpA3 [Kaufman, *Proc. Natl. Acad. Sci. USA*, 82:689-693 (1985)] by calcium phosphate coprecipitation and transfection.

Alternatively, the pEMC2B1 vector containing the meg-CSF gene in operative association with other plasmid sequences enabling expression thereof is introduced into DHFR-deficient CHO cells, DUKX-BII, by protoplast fusion and transfection. The meg-CSF gene and DHFR marker gene are both efficiently expressed when meg-CSF is introduced into pEMC2B1. The meg-CSF gene may be introduced into pMT2 as previously mentioned and the resultant vector used in place of pXM/meg-CSF and pAdA26SV(A)3.

DHFR expressing transformants are selected for growth in alpha media with dialyzed fetal calf serum. Transformants are checked for expression of meg-CSF by bioassay, immunoassay or RNA blotting and positive pools are subsequently selected for amplification by growth in increasing concentrations of MTX (sequential steps in 0.02, 0.2, 1.0 and 5uM MTX) as described in Kaufman et al., *Mol. Cell Biol.*, 5:1750 (1983). The amplified lines are cloned, and meg-CSF protein expression is monitored by the fibrin clot assay. meg-CSF expression is expected to increase with increasing levels of MTX resistance.

In any of the expression systems described above, the resulting cell lines can be further amplified by appropriate drug selection, resulting cell lines recloned and the level of expression assessed using the murine fibrin clot assay described above.

The meg-CSF expressing CHO cell lines can be adapted to growth in serum-free medium. Homogeneous meg-CSF can be isolated from conditioned medium from the cell line using methods familiar in the art, including techniques such as lectin-affinity chromatography, reverse phase HPLC, FPLC and the like.

Example 8 - Biological Activities of Human meg-CSF

The following assays were performed using the purified meg-CSF described in Example 1. The recombinant version of the molecule is expected to exhibit meg-CSF biological properties in these same assays or other assays.

A. Murine Fibrin Clot Assay

The meg-CSF obtained from Step 7 of the purification techniques of Example 1 was tested for activity in the megakaryocyte colony formation assay performed substantially as described in S. Kuriya et al, Exp. Hematol., 15:896-901 (1987). A fibrin clot was formed containing 2.5×10^5 mouse bone marrow cells in a 96-well plate. The diluted sample was layered around the clot and incubated for 6 days. Thereafter, cells were fixed and megakaryocytes were stained for acetylcholinesterase, a specific marker for murine megakaryocytes. A colony was defined as three or more megakaryocytes per unit area. Two types of megakaryocyte colonies were routinely observed: pure megakaryocyte colonies containing no additional cell types, and mixed megakaryocyte colonies containing additional non-megakaryocyte cell types.

The following control samples were included in every assay. A positive control was WEHI conditioned medium (murine IL-3), which produced between 7-25 (average 12) megakaryocyte colonies per clot, approximately 50% pure and 50% mixed megakaryocyte colonies. Another positive control was serum taken from lethally irradiated dogs at the nadir of the platelet count [see Mazur et al, Exp. Hematol., 13:1164-1172 (1985)], which produced between 6-22 (average 15) megakaryocyte colonies per clot, of which approximately 70% were pure and 30% were mixed megakaryocyte colonies. The negative control was Iscoves Medium, which produced 2-4 megakaryocyte colonies per clot.

In the assay, the meg-CSF has a specific activity of greater than approximately 5×10^7 dilution units/mg of protein. A unit of activity is defined as described in Example 3.

The major meg-CSF obtained from bone marrow transplant urine eluted from the S-Toyopearl cation exchange column chromatography step in the purification of Example 1 has been analyzed in this assay alone, together, and in combination with other cytokines. In the fibrin clot assay, it produced between 6-16 (average 13) megakaryocyte colonies, with 50-70% pure megakaryocyte colonies.

In each assay the samples were tested in duplicate and in three dilutions.

B. Human Plasma Clot meg-CSF Assay

The meg-CSF of this invention was also tested on an assay for human activity, the plasma clot meg-CSF assay described in E. Mazur et al, Blood, 57:277-286 (1981) with modifications. Non-adherent peripheral blood cells were isolated from Leukopacs and frozen in aliquots. The test sample was mixed with platelet-poor human AB plasma and 1.25×10^5 cells in 24-well plates and allowed to clot by the addition of calcium. After a 12 day incubation, megakaryocytes were identified using a monoclonal antibody directed against platelet glycoproteins IIb/IIIa and a horseradish peroxidase/anti-peroxidase chromogenic detection system. Recombinant human IL-3 [Genetics Institute, Inc.] was used as a positive control, producing 12-30 megakaryocyte colonies per clot with approximately 60% pure and 40% mixed megakaryocyte colonies. As in the murine assay, the aplastic dog serum was also used as a positive control, which produced between 5-10 megakaryocyte colonies per clot, of which approximately 50% were pure megakaryocyte colonies contained less than 10 cells, and 50% were mixed megakaryocyte colonies containing more than 40 megakaryocytes. The negative control was Alpha Medium, which produced 0-1 megakaryocyte colonies per clot.

The meg-CSF product from Step 8 of the above-described purification scheme may be active in this assay.

C. Murine meg-CSF Assay

An assay was performed on the meg-CSF from Step 7 of the purification according to P. J. Quensenberry et al, Blood, 65(1):214-217 (1985). In the assay, the meg-CSF stimulates the growth of acetylcholinesterase positive megakaryocyte colonies containing on average between 4-15 cells per colony. The sizes of the megakaryocytes are variable ranging from small immature cells to morphologically large mature cells.

D. Other assays

Several additional megakaryocyte assays using murine bone marrow cells were employed including the liquid acetylcholinesterase induction assay of Ishibashi et al, Blood, 69:1737-1741 (1987) and the liquid serotonin uptake assay of Vanucchi et al, exp. Hematol., 16:916-921 (1988).

Fractions were also routinely assayed in several factor dependent cell lines to screen for the presence of growth factors which alone or in combination might stimulate colony formation. The cell lines used were the human erythroleukemic cell line TF-1, the human megakaryoblastic cell line MO-7, the murine IL-6-dependent cell line T1165, and the murine IL-3-dependent cell line DA-1a.

The foregoing descriptions detail presently preferred embodiments of the invention. Numerous modifications and variations in practice of this invention are expected to occur to those skilled in the art. Such modifications and variations are encompassed within the following claims.

5 Claims

1. A gene encoding a megakaryocyte colony stimulating factor (meg-CSF) comprising a DNA sequence selected from
 - (a) the same or substantially the same DNA sequence as in Table I,
 - (b) the same or substantially the same DNA sequence as in Table II, or
 - (c) the same or substantially the same DNA sequence as in Table III.
2. The genomic KpnI-SnaBI DNA sequence encoding meg-CSF contained in ATCC 40857.
3. The genomic DNA sequence 18-5665 encoding meg-CSF contained in ATCC 40856.
4. A gene encoding a protein exhibiting properties of meg-CSF and comprising a DNA sequence:
 - (a) which differs from the DNA sequence of any one of claims 1 to 3 in codon sequence due to the degeneracy of the genetic code;
 - (b) which represents a fragment, allelic variation, analog or derivative of the DNA sequence of any one of claims 1 to 3; or
 - (c) which hybridizes with the DNA sequence of any one of claims 1 to 3 or section (a) or (b), above.
5. A gene encoding meg-CSF comprising all or a portion of the amino acid sequence from amino acid # 1 to amino acid # 182 of Table III.
6. An expression vector containing the gene or DNA sequence of any one of claims 1 to 5.
7. A host cell transformed with the vector of claim 6.
8. The host cell of claim 7, which is a mammalian cell, bacterial cell, yeast cell or insect cell.
9. The host cell of claim 8, which is a CHO cell or 3T3 cell.
10. A protein exhibiting properties of meg-CSF which is encoded by the gene or DNA sequence of any one of claims 1 to 5.
11. A protein exhibiting properties of meg-CSF which is obtainable by the steps of culturing in a suitable culture medium the host cell of any one of claims 7 to 9 and recovering said protein from said culture medium.
12. A process for producing meg-CSF comprising the steps of culturing the host cell of any one of claims 7 to 9 in a suitable culture medium and isolating said protein from said culture medium.
13. A pharmaceutical composition comprising meg-CSF or a fragment thereof in a pharmaceutically effective vehicle.
14. The composition according to claim 13 further comprising therapeutically effective amounts of at least one additional cytokine, hematopoietin, growth factor, thrombopoietin-like factor or antibody.
15. The composition according to claim 14 where said cytokine is selected from the group consisting of G-CSF, CSF-1, GM-CSF, IL-1, IL-3, IL-4, erythropoietin, IL-11, IL-6, TPO, M-CSF and IL-7.
16. The composition according to any one of claims 13 to 15 for treating bleeding disorders or platelet deficiencies.

A RESTRICTION MAP OF MEG-CSF GENOMIC CLONE

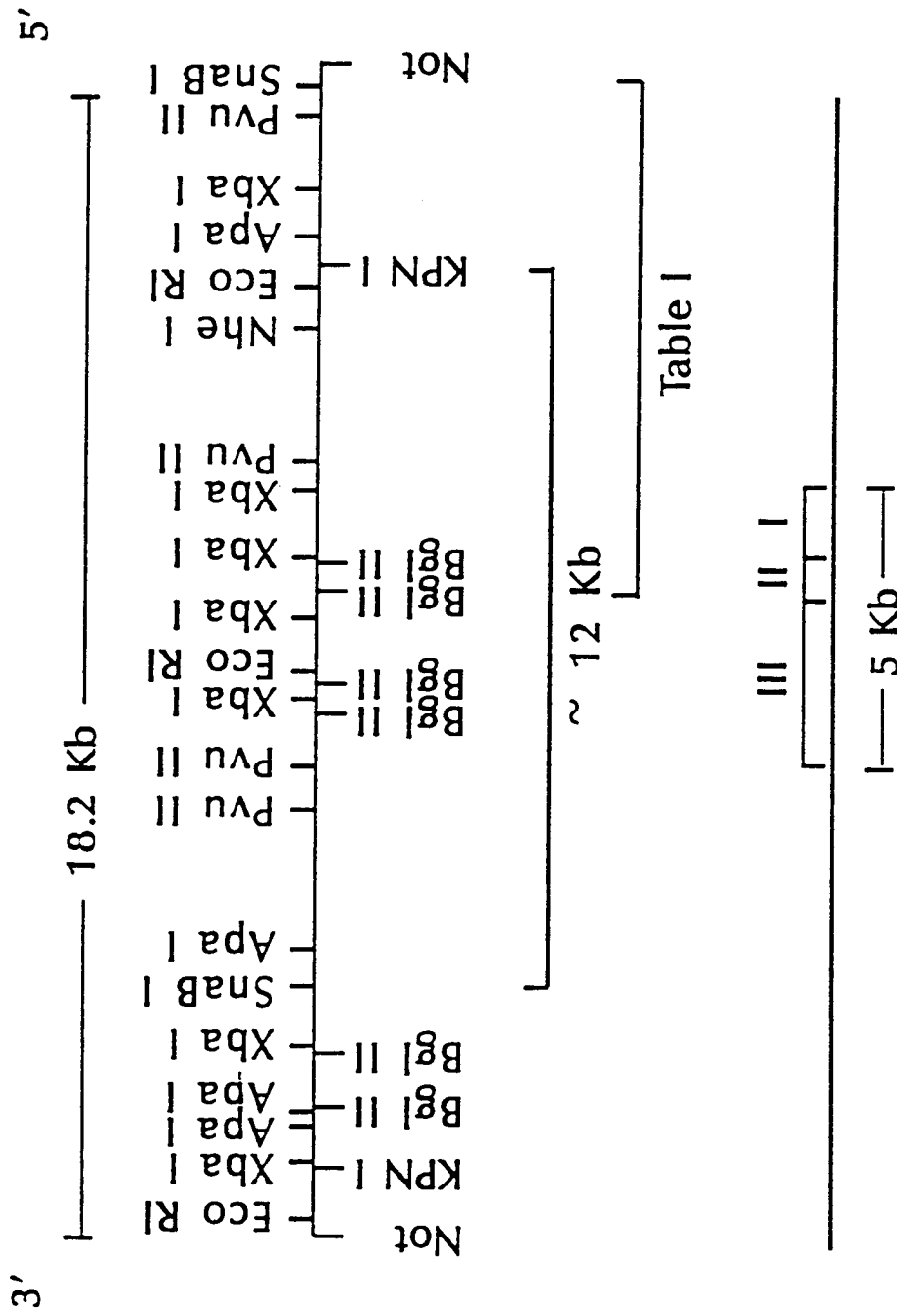


FIG. 1